# TOPICS IN FIXED EFFECTS FACTORIAL EXPERIMENTS
# COMPARING TESTS OF MAIN EFFECTS AND
# SHRINKING THE OLS ESTIMATOR TOWARD ADDITIVITY

BY

RICHARD OTIS LYNCH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1986

ACKNOWLEDGMENTS

I thank Dr. Ramon C. Littell for his help and advisement in this project. Also, Dr. Malay Ghosh made invaluable suggestions for which I am grateful.

A rough draft of this manuscript was typed using the word processor in the Statistics Department. I would like to thank Cindy Zimmerman for the dedication and hard work that was required to refine the rough draft into an acceptable form.

Two fundamental needs for a statistics researcher, computer resources and office space, were provided by IFAS Statistics.

My wife, Susan, has sacrificed more than can be written to allow me to complete this work. Though she expects nothing in return, I'll spend my life making it up to her.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ..... iii

LIST OF TABLES ..... vi

LIST OF FIGURES ..... viii

ABSTRACT ..... xi

CHAPTERS

1 INTRODUCTION ..... 1

1.1 Factorial Experiment ..... 2
1.1.1 Hypothesis Testing ..... 2
1.1.2 Parameter Estimation ..... 7
1.2 Questions Under Research ..... 10
1.2.1 Which Test Statistic Is Best? ..... 6
1.2.2 Should Biased Parameter Estimates Be Used? ..... 1

2 POWER COMPARISON OF THE WEIGHTED SQUARES OF MEANS METHOD AND THE METHOD OF FITTING CONSTANTS IN UNBALANCED ANOVA ..... 12

2.1 Introduction ..... 12
2.2 Power Function ..... 18
2.2.1 Bias Due to Interaction ..... 20
2.2.2 Power of Tests ..... 22
2.2.3 The 2x2 Case ..... 28
2.3 Conclusion ..... 30

3 BIASED ESTIMATORS IN LINEAR REGRESSION ..... 31

3.1 Introduction ..... 31
3.2 Criterion for Goodness in Estimators ..... 31
3.3 Preliminary Test Estimators ..... 34
3.4 Shrinkage Estimators ..... 36
3.4.1 James-Stein Estimators ..... 39
3.4.2 Empirical Bayes ..... 50

4 TWO-WAY FIXED EFFECTS FACTORIAL EXPERIMENT ..... 60

4.1 Application of the Estimators ..... 60
4.1.1 Conventional Strategy Estimator ..... 62

4.1.2 James-Stein Estimators ..... 61
4.1.3 Empirical Bayes ..... [illegible]
Examples ..... 65
4.2.1 Example 1 Nemetode Study ..... 66
4.2.2 Example 2 ..... 73
4.3 Monte Carlo Comparisons ..... 76
4.3.1 Simulation Description ..... 76
4.3.2 Conclusions ..... 83

CONFIDENCE INTERVALS ..... 97

5.1 Motivating Interval Estimation Strategies ..... 97
5.1.1 OLS Confidence Intervals ..... 97
5.1.2 Restricted Least Squares Confidence Intervals ..... 98
5.1.3 Preliminary Test Confidence Interval ..... 98
5.1.4 James-Stein Confidence Interval ..... 99
5.2 Example Nemetode Study ..... 103
5.3 Monte Carlo Comparison ..... 109
5.3.1 Simulation Description ..... 109
5.3.2 Simulation Results ..... 110
5.3.3 Conclusions ..... 142

CONCLUSIONS ..... 143

6.1 Weighted Squares of Means vs Method of Fitting Constant ..... 143
6.2 Compromise Estimators in Factorial Experiments ..... 144
6.3 Confidence Intervals ..... 145
6.3.1 OLS Confidence Interval ..... 145
6.3.2 Restricted Confidence Intervals ..... 146
6.3.3 Preliminary Test Confidence Interval ..... 146
6.3.4 James-Stein Confidence Interval ..... 147
6.4 Recommendations ..... 149

APPENDICES

A GENERATING REPRESENTATIVE PATTERNS OF INTERACTION FOR SIMULATION OF A TWO-WAY FACTORIAL EXPERIMENT ..... 150

B COMPUTER PROGRAMS FOR COMPUTING BIASED ESTIMATES OF CELL MEANS ..... 156

BIBLIOGRAPHY ..... 169

BIOGRAPHICAL SKETCH ..... 173

## LIST OF TABLES


| Table | | Page |
|---|---|---|
| 2.1 | Noncentrality Parameters ($\lambda$) and Rejection Probabilities ($Prob_\lambda$) for Various Tests and Selected Sample Sizes | 24 |
| 4.1 | Summary Information Required to Compute Estimators | 65 |
| 4.2 | Square Root of Number of Observed Nematodes for Each Nematocide-Fertilizer Combination | 67 |
| 4.3 | ANOV Table for Nematode Experiment | 68 |
| 4.4 | Predicted Cell Means and Marginal Means Using Various Estimators | 69 |
| 4.5 | Summary Information for Example 2 | 73 |
| 4.6 | Predicted Cell Means and Marginal Means Using Various Estimators | 74 |
| 4.7 | Cell Sizes Considered in the Simulations | 79 |
| 4.8 | Patterns of Interaction Used in Simulation | 81 |
| 4.9 | Risk Increase ($D_1$) of Restricted Estimator over OLS Estimator, Situation 1 | 84 |
| 4.10 | Risk Increase ($D_1$) of PTE over OLS Estimator. Situation 1 | 85 |
| 4.11 | Risk Increase ($D_1$) of James-Stein + Estimator over OLS Estimator, Situation 1, Table 4.7 | 86 |
| 4.12 | Risk Increase ($D_1$) of Empirical Bayes + Estimator over OLS Estimator, Situation 1, Table 4.7 | 87 |
| 4.13 | Risk Increase ($D_2$) of Restricted Estimator over OLS Estimator, Situation 1, Table 4.7 | 88 |
| 4.14 | Risk Increase ($D_2$) of PTE over OLS Estimator. Situation 1, Table 4.7 | 89 |

4.15 Risk Increase ($D_2$) of James-Stein + Estimator over OLS Estimator, Situation 1, Table 4.7 ..... 90

4.16 Risk Increase ($D_2$) of Empirical Bayes + Estimator over OLS Estimator, Situation 1, Table 4.7 ..... 91

4.17 Risk Increase ($D_1$) of the Biased Estimators over OLS Estimator, Situation 2, Table 4.7 ..... 92

5.1 Listing of $(I-MM)S^{-1}$ for Nematode Study ..... 105

5.2 Listing of D for the Nematode Study ..... 107

5.3 Confidence Intervals for Nematode Study ..... 108

5.4 Cell Sizes Considered in the Simulations ..... 110

5.5 Numbering of Cells in 3×3 Layout ..... 110

A.1 Patterns of Interaction for the Simulation of the Unbalanced Case, p=79 ..... 155

LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 2.1 | Harmonic Mean of $n_{11}$ and $n_{21}$ vs Harmonic Mean of $n_{12}$ and $n_{22}$ | 21 |
| 2.2 | Type 1 Error Rate of $F_A(11)$ | 23 |
| 2.3 | Power of $F_A(11)$ | 25 |
| 2.4 | Power of $F_A(111)$ | 26 |
| 2.5 | Rejection Probability of $F_A(11)$ | [illegible] |
| 2.6 | Power of Test for Interaction | [illegible] |
| 3.1 | Risk of OLS and James-Stein Estimators | [illegible] |
| 4.1 | Geometric Interpretation of James-Stein Estimate of Cell Means Using Nematode Example | [illegible] |
| 4.2 | Geometric Interpretation of James-Stein Estimate of Marginal Means Using Nematode Study | 71 |
| 4.3 | Impact on Compromise Estimators on Interpretation of Results in Nematode Example | 72 |
| 4.4 | Geometric Interpretation of James-Stein Estimate of Cell Means Using Example 2 Data | 75 |
| 4.5 | Geometric Interpretation of James-Stein Estimate of Marginal Means Using Example 2 | 76 |
| 4.6 | Impact of Compromise Estimators on Interpretation of Results in Example 2 | 77 |
| 4.7 | Increased Risk $D_1$ of Biased Estimates of the Cell Means over the OLS Estimate of the Cell Means, Situation 1, Table 4.7 | 93 |
| 4.8 | Increased Risk $D_2$ of Biased Estimates of the Marginal Means over the OLS Estimate of the Marginal Means, Situation 1, Table 4.7 | 94 |

4.8 Increased h[illegible] of Biased Estimate of [illegible] Cell Means over the OLS [illegible]are of the Cell Means, Situation 2, Table 4.7 ..... 95

5.1 Estimated Coverage Probabilities and Average Half Widths for Cell 1 in Situation 1, Table 5.4 ..... 114

5.2 Estimated Coverage Probabilities and Average Half Widths for Cell 2 in Situation 1, Table 5.4 ..... 116

5.3 Estimated Coverage Probabilities and Average Half Widths for Cell 3 in Situation 1, Table 5.4 ..... 118

5.4 Estimated Coverage Probabilities and Average Half Widths for Cell 4 in Situation 1, Table 5.4 ..... 120

5.5 Estimated Coverage Probabilities and Average Half Widths for Cell 5 in Situation 1, Table 5.4 ..... 122

5.6 Estimated Coverage Probabilities and Average Half Widths for Cell 6 in Situation 1, Table 5.4 ..... 124

5.7 Estimated Coverage Probabilities and Average Half Widths for Cell 7 in Situation 1, Table 5.4 ..... 126

5.8 Estimated Coverage Probabilities and Average Half Widths for Cell 8 in Situation 1, Table 5.4 ..... 128

5.9 Estimated Coverage Probabilities and Average Half Widths for Cell 9 in Situation 1, Table 5.4 ..... 130

5.10 Estimated Coverage Probabilities and Average Half Widths for the A=1 Marginal Mean in Situation 1, Table 5.4 ..... 132

5.11 Estimated Coverage Probabilities and Average Half Widths for the A=2 Marginal Mean in Situation 1, Table 5.4 ..... 134

5.12 Estimated Coverage Probabilities and Average Half Widths for the A=3 Marginal Mean in Situation 1, Table 5.4 ..... 136

5.13 Estimated Coverage Probabilities and Average Half Widths for Situation 2, Table 5.4 ..... 139

5.14 Estimated Coverage Probabilities and Average Half Widths for the Various Confidence Intervals of Cell Mean 6 in the Unbalanced Situation ..... 140

[illegible] Estimated Coverage Probabilities and Average Half Widths for the Various Confidence Intervals of A=2 Marginal Means in the Unbalanced Situation [illegible]

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TOPICS IN FIXED EFFECTS FACTORIAL EXPERIMENTS
COMPARING TESTS OF MAIN EFFECTS AND
SHRINKING THE OLS ESTIMATOR TOWARD ADDITIVITY

BY

RICHARD OTIS LYNCH

December 1986

Chairman: Ramon C. Littell
Major Department: Statistics

Two methods of testing for significance of a main effect in a fixed effects two-way factorial experiment are the method of fitting constants and the weighted squares of means method. In the event that the cell sample sizes are unequal (unbalanced), the two methods can give substantially different results. The method of fitting constants exhibits higher power, as a result of assuming additivity (no interaction) between the two main effects. In the event that interaction actually exists, the method of fitting constants can have a larger type I error rate than prescribed, depending upon the pattern of cell sample sizes. A recommendation is made against the method of fitting constants.

Two extreme methods for estimating the vector of cell means, or the vector of marginal means, are the restricted least squares method assuming additivity (restricted estimator) and the ordinary least squares method (OLS estimator) that does not assume additivity. A

Preliminary test estimator that allows the outcome of a test for interaction significance to dictate the choice between the restricted and OLS estimators is often used by practicing statisticians. The James-Stein and Empirical Bayes methods of compromising between the restricted and OLS estimators are developed for $(a-1)(b-1) \geq 3$. Simulation experiments demonstrate the advantages of the James-Stein and Empirical Bayes estimators over the restricted, OLS and preliminary test estimators in terms of closeness to true cell means and true marginal means. A recommendation is made for the use of the James-Stein estimator rather than the preliminary test estimator.

Confidence intervals for linear combinations of cell means are often calculated by the applied statisticians. Four interval estimation strategies are described that are motivated by the OLS, restricted, preliminary test and James-Stein estimators. Using the simulation of specific 3×3 factorial experiments, coverage probabilities and average interval half widths are estimated for the four strategies. Other confidence intervals are found to have smaller average half width than the OLS confidence interval, at the expense of possible coverage probability reduction. The performance of the James-Stein confidence interval is found to be slightly more desirable than the preliminary test confidence interval.

# CHAPTER 1
# INTRODUCTION

[illegible] addresses two topics of importance in the [illegible] analyzing data from a two-way factorial experiment. The [illegible] is the testing for significance of main effects. The second topic is the estimation of cell means and marginal means.

The objective in chapter 2 is to compare power and type I error rates for two methods of testing for significance of main effects. The method of fitting constants tests for differences in main effects, assuming additivity (no interaction). The weighted squares of means method tests for differences in main effects, unconditionally. In the case that unequal replication of the factorial combinations occurs (unbalance), the two methods may lead to conflicting conclusions. The method of fitting constants is a more powerful method of testing if additivity exists. If nonadditivity occurs, the method of fitting constants may have larger than expected type I error rate. For the 2×2 case and several patterns of unbalance, the power and type I error rate of the two methods are computed.

The main objective for the remainder of this study is to determine the practical value of using compromise estimators for main effects and marginal means. Specifically, compromises between the estimator that assumes additivity and the estimator that allows nonadditivity are considered. The types of compromise estimators in question are the James-Stein and Empirical Bayes shrinkage estimators. These estimators are

compared with the ordinary least squares and preliminary test estimators in a series of Monte Carlo experiments. The criteria for comparison are how close the estimators tend to be to the true cell means (and marginal means), as well as size and coverage probability of confidence intervals based on the various estimators. Implicit in this objective is the requirement that these estimators be stated in a form that is understandable to the average applied statistician. A large volume of this report involves development of the form of these estimators for the situation of compromising between an ordinary least squares estimator and a restricted least squares estimator. The estimation problem in the two way factorial experiment is treated as a specific application of these estimators. Examples are included to demonstrate the computations required to make use of compromise estimators.

## 1-1 Factorial Experiment

A factorial experiment is an experiment in which all combinations of the levels of the factors being studied are applied to experimental units. Some of the early work done on the statistical methodology associated with factorial experiments appears in Yates (1935). Since that time, the factorial experiment has come to be a cornerstone in text books on experimental design and analysis (see Cochran and Cox (1957), Steel and Torrie (1980), and Kempthorne (1952)). This experiment may be laid out in one of many possible designs (e.g. completely randomized, randomized block, or split plot); in this discourse, unless otherwise noted, the experiment is assumed to be laid out in a completely randomized design.

A factor is either a fixed effect or a random effect. The factor F is a fixed effect if the levels of F being studied in the experiment comprise the entire set of levels of F, about which inference shall be made. In contrast, the factor R is a random effect if the levels of R being studied in the experiment are a subset taken randomly from a larger population of levels of R, and it is the larger population of levels about which inference shall be made.

A <u>fixed effects factorial experiment</u> is one in which all factors being applied are <u>fixed effects</u>. A <u>random effects factorial experiment</u> is one in which all factors being applied are <u>random effects</u>. A <u>mixed model factorial experiment</u> is one in which some of the factors being applied are <u>fixed effects</u>, and some of the factors being applied are <u>random effects</u>.

Any number of factors can be studied in a factorial experiment. In this discourse the case of two factors is studied. This case will be referred to as a two-way factorial experiment.

The two-way fixed effects factorial experiment laid out in a completely randomized design is described as follows.

The effect of two factors A and B on a stochastic response, y, needs to be determined. The factor A takes on multiple qualitative levels, indexed $i=1,2,\ldots,a$. The factor B takes on multiple qualitative levels, indexed $j=1,2,\ldots,b$. The factor level combination $(i,j)$ is referred to as treatment $(i,j)$, for $i=1,2,\ldots,a$ and $j=1,2,\ldots,b$. There are $n_{ij}$ experimental units subjected to treatment $(i,j)$. Unless

[illegible]ted, we will assume that $n_{ij}$ [illegible] $i=1,2,\ldots,a$ and $j=1,2,\ldots,b$. For all of the $N=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}$ experimental units, the response, $y$, is measured. The experimental units are assigned to the treatments randomly.

The following linear model, known as the <u>factor effects model</u>, is used to describe $y$ as a function of the treatments

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad , \qquad (1.1)$$

where $i=1,2,\ldots,a$, $j=1,2,\ldots,b$ and $k=1,2,\ldots,n_{ij}$. To obtain estimability of the parameters, it will be assumed in this work that

$$\sum_{i=1}^{a}\alpha_i = \sum_{j=1}^{b}\beta_j = \sum_{i=1}^{a}\gamma_{ij} = \sum_{j=1}^{b}\gamma_{ij} = 0.$$

This may be written in a shorthand notation as

$$\alpha_{\cdot} = \beta_{\cdot} = \gamma_{i\cdot} = \gamma_{\cdot j} = 0.$$

Further, it will be assumed that $\epsilon_{ijk}$, $i=1,2,\ldots,a$, $j=1,2,\ldots,b$ and $k=1,2,\ldots,n_{ij}$ are independently and identically distributed normal $(0,\sigma^2)$ random variables. This is denoted $\epsilon_{ijk} \sim i.i.d.\ N(0,\sigma^2)$.

The factor effects model can be reformulated as the <u>cell means model</u>, by taking $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$. The cell means model is

[illegible] $j=1,\ldots,b$ and $k=1,\ldots,n_{ij}$ [illegible] the cell means model is written:

$$\underline{y}=X\underline{\mu}+\underline{\epsilon},\ \epsilon\sim N(0,\sigma^2 I) \qquad (1.1.1)$$

where

$$\underline{y}=(y_{111},y_{112},\ldots,y_{11n_{11}},y_{121},\ldots,y_{1b1},\ldots,y_{1bn_{1b}},\ldots,y_{a11},\ldots,y_{abn_{ab}})'$$

$$X=\begin{pmatrix} \underline{1}_{n_{11}} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{1}_{n_{12}} & \cdots & \underline{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{1}_{n_{ab}} \end{pmatrix}$$

$$\underline{\mu}=(\mu_{11},\mu_{12},\ldots,\mu_{1b},\mu_{21},\ldots,\mu_{a1},\ldots,\mu_{ab})',$$

$$\underline{\epsilon}=(\epsilon_{111},\epsilon_{112},\ldots,\epsilon_{11n_{11}},\ldots,\epsilon_{ab1},\ldots,\epsilon_{abn_{ab}})'.$$

Note that $\underline{1}_m$ denotes a 1xm vector of ones. For details on working with linear models through matrix algebra, refer to Searle (1971)

1.1.1 Hypothesis Testing

For almost every situation in which linear models such as (1.1.1) and (1.1.2) are used there will be certain linear hypotheses about the

parameter vector that are of interest. Referring to the parameters in (1.1.1), some commonly tested linear hypotheses are

1. $H_0$: $\gamma_{ij}=0$ for all $i=1,2,\ldots,a$ and $j=1,2,\ldots,b$;
   $H_1$: Not $H_0$.
2. $H_0$: $\gamma_{ij}=0$ and $\delta_j=0$ for all $i=1,2,\ldots,a$ and $j=1,2,\ldots,b$.
   $H_1$: Not $H_0$.
3. $H_0$: $\alpha_i=0$ for all $i=1,2,\ldots,a$;
   $H_1$: Not $H_0$.

A linear hypothesis about parameters in a linear model may be stated in terms of a restriction matrix. For example, in model (1.1.2) the linear hypothesis $H_0$: $\mu_{11}=\mu_{12}=\mu_{13}$ may be written

$$H_0: R\underline{\mu} = \underline{0},$$

where $\underline{\mu}$ is as described in (1.1.3) and

$$R = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}.$$

In this discourse the restriction matrices considered are all of full row rank, and are $J \times K$, where $K$ is the number of elements in the parameter vector and $J \leq K$.

To test a linear hypothesis such as

$$H_0 \quad R\underline{\mu}=\underline{r} \text{ vs } H_1 \quad \text{Not } H_0,$$

the F-test statistic is computed. The F-test statistic is

$$F = \frac{(R\hat{\underline{\mu}}-\underline{r})'(R(X'X)^{-1}R')^{-1}(R\hat{\underline{\mu}}-\underline{r})}{s^2 \text{ rank}(R)} , \qquad (1.1.4)$$

where $\hat{\underline{\mu}} = (X'X)^{-1}X'\underline{y}$ , $s^2 = (\underline{y}-X\hat{\underline{\mu}})'(\underline{y}-X\hat{\underline{\mu}})/(N-K)$ and rank (R) is the rank of R. Recall that $N = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}$ , and let K be the number of elements in $\underline{\mu}$. Then, assuming $H_0$ is true, F has the central F distribution, with degrees of freedom rank(R) and N-K. In general F has the noncentral F distribution with degrees of freedom rank(R) and N-K, and noncentrality parameter

$$\lambda = (R\underline{\mu}-\underline{r})'(R(X'X)^{-1}R')^{-1}(R\underline{\mu}-\underline{r})/2\sigma^2 . \qquad (1.1.5)$$

1.1.2 Parameter Estimation

For the linear model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim N(\underline{0}, I\sigma^2)$, the OLS estimate of $\underline{\beta}$ is $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$. If the estimate is restricted to the subspace $\{R\underline{\beta}=\underline{r}\}$, then the restricted estimate of $\underline{\beta}$ is $\tilde{\underline{\beta}} = (I-MR)\hat{\underline{\beta}} + M\underline{r}$, where $M = (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}$. Derivation of the restricted estimate is given in Searle (1971, p. 113).

As a specific case consider the model stated in (1.1.3). The unrestricted least squares estimate of $\underline{\mu}$ is

$$E(\hat{\underline{\mu}}) = \underline{\mu} \text{ and } Var(\hat{\underline{\mu}}) = (X'X)^{-1}\sigma^2 = Diag(n_{11}, \ldots, n_{ab})^{-1}\sigma^2$$

For the sake of brevity $(X'X)^{-1}$ will be denoted $S^{-1}$ in the rest of this work.

If it is assumed that there is no interaction between A and B, then a restricted least squares estimate for $\underline{\mu}$ can be formulated. First, the assumption of no interaction can be stated as

$$R\underline{\mu} = \underline{0},$$

where $R = (-\underline{1} \quad I_{(a-1)}) \otimes (-\underline{1} \quad I_{(b-1)})$. The symbol $\otimes$ denotes the left direct product, as described by Graybill (1969). The rank of R is $(a-1)(b-1)$. The restricted least squares estimate of $\underline{\mu}$ assuming $R\underline{\mu} = \underline{0}$ is

$$\bar{\underline{\mu}} = \hat{\underline{\mu}} - S^{-1}R'(RS^{-1}R')^{-1}(R\hat{\underline{\mu}}) \qquad (1.1.7)$$

Following the notation of Hook et. al. (1973), let

When $\underline{\mu}$ can [illegible]

$$\tilde{\underline{\mu}} = \hat{\underline{\mu}} - MR\,\hat{\underline{\mu}} \qquad (\ldots)$$

Some properties of $\tilde{\underline{\mu}}$ are

$$E(\tilde{\underline{\mu}}) = \underline{\mu} - MR\,\underline{\mu} \quad (= \underline{\mu} \text{ assuming } R\underline{\mu} = \underline{0}) \text{ and}$$

and

$$Var(\tilde{\underline{\mu}}) = (S^{-1} - MRS^{-1})\sigma^2 .$$

Notice that if indeed $R\underline{\mu} = \underline{0}$, both $\hat{\underline{\mu}}$ and $\tilde{\underline{\mu}}$ are unbiased estimators of $\underline{\mu}$, but $\tilde{\underline{\mu}}$ has smaller variance. By smaller variance it is meant that for any linear combination of the cell means $\underline{a}'\underline{\mu}$, $Var(\underline{a}'\tilde{\underline{\mu}}) \le Var(\underline{a}'\hat{\underline{\mu}})$, with the inequality strictly holding for at least one $\underline{a}'$. More succinctly, $Var(\hat{\underline{\mu}}) - Var(\tilde{\underline{\mu}})$ is a nonnegative definite matrix.

Often, the marginal means are of interest in a factorial experiment. The marginal means of the effect A for the two way factorial experiment are

$$\mu_{i.} = \frac{1}{b}\sum_{j=1}^{b} \mu_{ij} \ , \ i=1,2,\ldots,a \ .$$

The marginal means of the effect B for the two way factorial experiment are

$$\mu_{.j} = \frac{1}{a} \sum_{i=1}^{a} \mu_{ij} \qquad j=1,2,\ldots,b.$$

These marginal means are then combined into the vector $\underline{\mu}_M$ where $\underline{\mu}_M$ in vector notation is

$$\underline{\mu}_M = \begin{pmatrix} a^{-1} ( I_a \otimes \underline{1}_b' ) \\ a^{-1} ( \underline{1}_a' \otimes I_b ) \end{pmatrix} \underline{\mu} \qquad (1.1.4)$$

The symbol $\otimes$ denotes the left direct product, as described by Graybill (1969).

## 1.2 Questions Under Research

The purpose of this study is to seek answers to two questions facing the applied statistician analyzing the results of a two-way fixed effects factorial experiment. The questions under research are now described.

### 1.2.1 Which Test Statistic Is Best?

In the two-way fixed effects factorial experiment and for the model given in (1.1.1) the following hypothesis is often tested:

$$H_0: \alpha_i = 0 \ , \ i=1,2,\ldots,a. \quad \text{vs} \quad H_1: \text{Not } H_0. \qquad (1.2.1)$$

This test can be made assuming $Y_{ij}=0$ $i=1,2,\ldots,a$ and $j=1,2,\ldots,b$. Such a conditional test arises from the "method of fitting constants." The computer procedure SAS/GLM calls this conditional test the "Type II F-test." The hypothesis in (1.2.1) can be tested unconditionally using

the techniques described in section 1.1. This unconditional test arises from the "weighted squares of means method". The computer procedure SAS/GLM calls the unconditional test the "Type III F-test." In chapter 2 these two test statistics are compared to determine which test statistic is better.

1.2.2 Should Biased Parameter Estimates Be Used?

In section 1.1 two estimators of the parameter vector $\underline{\mu}$ are given. The OLS estimator in (1.1.6) is unbiased, while the restricted estimator in (1.1.8) is biased, in general. There are numerous estimation strategies that attempt to compromise between (1.1.6) and (1.1.8). These compromise estimators are themselves biased. In chapter 3, several compromise estimators are discussed for a general linear model. In chapter 4, these compromise estimators are applied to the two-way fixed effects factorial experiment, and their performance is compared. Finally, in chapter 5, confidence intervals associated with compromise estimators are developed and applied.

## POWER COMPARISON OF THE WEIGHTED SQUARES OF MEANS [illegible] THE METHOD OF FITTING CONSTANTS IN UNBALANCED A[illegible]

### 2.1 Introduction

This chapter deals with the comparison of the weighted squares of means method and the method of fitting constants in unbalanced ANOVA. Specifically, the two-way fixed effects factorial experiment with factors A and B, both fixed, with a levels of A and b levels of B, is addressed. Recall from the description of a factorial experiment in chapter 1 that there are $n_{ij}$ observation $y_{ijk}$, $k=1,\ldots,n_{ij} > 0$, from the population corresponding to levels i of A and j of B, $i=1,\ldots,a$ and $j=1,\ldots,b$. Using the cell means model as in (1.1.2), $\mu_{ij}$ denotes the mean of the ij population. Assume $V(y_{ijk}) = \sigma^2$ independent of i and j, and that all $y_{ijk}$ are mutually independent and normally distributed. Reiterating, the cell means model is $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, where the $\varepsilon_{ijk}$ are independent and normally distributed with variance $\sigma^2$. Defining

$\mu = \bar{\mu}_{..}$, $\alpha_i = \bar{\mu}_{i.} - \bar{\mu}_{..}$, $\beta_j = \bar{\mu}_{.j} - \bar{\mu}_{..}$, and

$\gamma_{ij} = \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}$ the factor effects model can be derived from the cell means model. Recall that the factor effects model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad (2.1.1)$$

where $\alpha = \beta_j = \gamma_{i.} = \gamma_{.j} = 0$, $i=1,\ldots,a$, $j=1,\ldots,b$, $k=1,\ldots,n_{ij}$. The

[illegible] $\alpha_i$, $\beta_j$ and $\gamma_{ij}$ measures A main effect, B main effect, and A*B interaction, respectively.

Suppose that it is of interest to test whether there is a main effect of factor A. Specifically, it is desired to test $H_0^A$: $\alpha_1 = \ldots = \alpha_a = 0$, or equivalently $H_0^A$: $\bar{\mu}_{1\cdot} = \ldots = \bar{\mu}_{a\cdot}$. Note that this can be considered a main effect test in the possible presence of interaction. In many applications this may be irrelevant; it may be more meaningful to analyze simple effects. There are, however, situations in which main effects are important even with interaction, e.g., if b is too large to feasibly separate all simple effects, if practical considerations require a single overall judgement, or if interaction is small relative to main effects.

Several papers have addressed the topic of hypothesis testing in unbalanced data, notably Yates (1934) in the early years. Textbook discussions are offered by Bancroft (1968), Searle (1971), Steel and Torrie (1960), (1980), and elsewhere. The topic has been of considerable recent interest due largely to the variety of computational procedures available in computer packages such as SAS. Survey papers are given by Speed, Hocking, and Hackney (1978) and Searle, Speed, and Henderson (1981). A primary focus of these writings has concerned the concept of the "hypotheses tested" by the F-tests based on the various sums of squares in an analysis of variance. The term "hypotheses tested" can be explained as follows: An F-statistic in a fixed effects model has a noncentral F distribution. The "hypotheses tested" by an F-test are a set of equations, linear in the parameters, that are true if and only if the noncentrality parameter is equal to

zero. Thus the F-statistic has a central chi-square distribution when the "hypothesis tested" is true. Hence a critical value $F_\alpha$ can be determined so that the rejection probability is equal to the prescribed $\alpha$ when the hypothesis tested is true and greater than $\alpha$ when the hypothesis tested is false. Put another way, an F-statistic tests $H_0$ if the F-test is unbiased for $H_0$. Conversely, an F-statistic does not test $H_0$ if it is biased for $H_0$; that is, if the rejection probability can exceed the prescribed $\alpha$ when $H_0$ is true.

The intent of this chapter is to weigh the merits of two methods of testing $H_0^A$. These methods are the "method of fitting constants" and the "weighted squares of means". (See Searle et al., 1981.) The method of fitting constants tests $H_0^A$ assuming the $\gamma_{ij}$'s are all zero. The weighted squares of means method does not make that assumption. The computer package SAS presents test statistics for both methods in PROC GLM. SAS uses the names "Type II" and "Type III" for the test statistics corresponding to the method of fitting constants and the weighted squares of means method, respectively. Henceforth, in this work the terminology of SAS will be used in the interest of brevity.

Speed, Hocking, and Hackney (1978) presented the linear functions of cell means that are tested by the Types II and III F-tests in SAS, as well as several other types. In particular, they stated that the Type III F-test does test $H_0^A: \bar{\mu}_{1.} = \cdots = \bar{\mu}_{a.}$, but that the Type II F-test does not test $H_0^A$. This same point is made in somewhat different terminology by Bancroft (1968) and Searle (1971).

A more detailed discussion of the types II and III SS's is now given. To do so, consider the four models

$$\alpha_i = \mu + a_i + \delta_{i(j)} + \varepsilon_{ijk}$$

$$a_i = \delta_{i(j)} = 0$$

$$y_{ijk} = \mu + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$\beta_j = \gamma_{ij} = 0$$

$$y_{ijk} = \mu + \beta_j + \varepsilon_{ijk}$$

$$\beta_j = 0$$

Let $SS_{E,\ell}$ and $df_{E,\ell}$ denote, respectively, the residual sums of squares and degrees of freedom for the four models, $\ell = 1, 2, 3, 4$. The Types II and III SS's for factor A are then expressible in the usual "R" notation, as

$$SS_A(II) = R(\alpha|\mu,\beta) = SS_{E,4} - SS_{E,2} \qquad (2.1.3)$$

and

$$SS_A(III) = R(\alpha|\mu,\beta,\gamma) = SS_{E,3} - SS_{E,1} \qquad (2.1.4)$$

(Note: Our "R" notation here corresponds to the Σ-restricted reduction $R^*$ notation of Searle (1981).)

If all $n_{ij}$ are equal, then $SS_A$ (II) = $SS_A$ (III); otherwise inequality holds, generally.

Referring to the R notation, the following guidelines are now stated:

i) In the absence of interaction, use Type II

ii) In the presence of interaction and assuming $H_0^A$ is of interest, use Type III.

The choice between Types II and III, therefore, hinges on the presence or absence of interaction. In application, however, the data analyst rarely knows whether interaction is present. In fact, interaction is probably always present, at least in a minute amount. A common practice, which is suggested by Bancroft (1968), is to first of all test for significance of interaction at, say, the $\alpha = .25$ level. If interaction is not significant, use the Type II test. If interaction is significant (and main effects are still of interest), use the Type III test.

As noted, if interaction is present, then the type II test is biased for testing $H_0^A : \bar{\mu}_{1.} = \cdots = \bar{\mu}_{a.}$ in the sense that the probability of rejecting $H_0^A$ will be larger than the prescribed $\alpha$ i.e. the type I error rate is inflated, and amount of inflation depends on the magnitude of the interaction. For this reason it is said that the type II F does not test $H_0^A$. Further, the probability of rejecting $H_0^A$ when $H_0^A$ is false (the power of the test) depends on the interaction, and this dependency can result in either an increase or a decrease in power.

The type III test always gives an unbiased test of $H_0^A$, in that the rejection probability is equal to $\alpha$ under $H_0^A$ regardless of the amount of interaction. Therefore, it is said that the type III F does test $H_0^A$. However, as noted e.g. by Steel and Torrie (1960) and Bancroft (1968),

If interaction is absent, then the type III test is [illegible]. Specifically, the power of the type III test is less than the power of the type II test in the absence of interaction, due to the superfluous estimation of zero-valued interaction parameters. The practical issue when small amounts of interaction are known or suspected to exist becomes a trade-off between a small bias in the type II test and a small power deficiency in the type III test.

Whereas the issue of what hypothesis is being tested by a given SS has been extensively discussed, relatively much less has been written about the amount of bias (error rate inflation) in the type II test when interaction is present. Also, relatively little has been reported on the degree of inefficiency of the type III test when interaction is absent. Gosslee and Lucas (1965) make limited power computations for both the types II and III tests. Overall, Lee, and Hardick (1981) and Cramer and Appelbaum (1980) debate related issues primarily regarding estimation.

The purpose of this chapter is to assess rejection probabilities of the types II and III F-tests. Specifically, computations are made for:

i. The amount of inflation in the type I error rate of the type II F-test in the presence of interaction.

ii. The loss in power (inefficiency) of the type III F-test in the absence of interaction.

iii. The effect that interaction has on the power of the type II F-test

In order to reduce the problem to a manageable size, discussion is focused primarily on the 2×2 case. In such case, $\alpha_2 = -\alpha_1$, $\beta_2 = -\beta_1$, and $\gamma_{22} = -\gamma_{21} = -\gamma_{12} = \gamma_{11}$

## 2.2 Power Function

In the notation above, the two F-tests considered for an A main effect are

$$F_A(II) = MS_A(II)/MS_E \qquad (2.2.1)$$

and

$$F_A(III) = MS_A(III)/MS_E \qquad (2.2.2)$$

where $SS_E = \sum_{ijk} (y_{ijk} - \bar{y}_{ij})^2$. Note that, like GLM, the $SS_{A*B}$ is not pooled with $SS_E$ in $F_A(II)$.

The sums of squares $SS_A(II)$ and $SS_A(III)$ can be computed according to the "reduction" method in (2.1.3) and (2.1.4). A more direct method for $SS_A(III)$ is

$$SS_A(III) = \sum_{i=1}^{a} H_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 , \qquad (2.2.3)$$

where

$$H_i = b^2 \left( \sum_{j=1}^{b} (1/n_{ij}) \right)^{-1},$$

$$\bar{y}_{i\cdot} = \left( \sum_{j=1}^{a} y_{ij} \right)/a$$

and

$$\bar{y}_{\cdot\cdot} = \left( \sum_{i=1}^{a} H_i \bar{y}_{i\cdot} \right) / \sum_{i=1}^{a} H_i$$

In the 2×2 case that is considered here, the SS's can be obtained from the contrasts

$$C_A(II) = h_1(\bar{y}_{11.} - \bar{y}_{21.}) + h_2(\bar{y}_{12.} - \bar{y}_{22.})$$

and

$$C_A(III) = \bar{y}_{11.} - \bar{y}_{21.} + \bar{y}_{12.} - \bar{y}_{22.}$$

where $h_j = 2(n_{1j}^{-1} + n_{2j}^{-1})^{-1}$ is the harmonic mean of $n_{1j}$ and $n_{2j}$. Specifically,

$$SS_A(II) = \frac{[h_1(\bar{y}_{11.} - \bar{y}_{21.}) + h_2(\bar{y}_{12.} - \bar{y}_{22.})]^2}{2(h_1 + h_2)} \qquad (2.2.4)$$

and

$$SS_A(III) = \frac{[\bar{y}_{11.} - \bar{y}_{21.} + \bar{y}_{12.} - \bar{y}_{22.}]^2}{2(h_1^{-1} + h_2^{-1})} \qquad (2.2.5)$$

Note that proportional data structure $n_{11}n_{22} = n_{12}n_{21}$ is not sufficient for $SS_A(II) = SS_A(III)$ to hold; a common misconception.

The F statistics corresponding to $SS_A(II)$ and $SS_A(III)$ have non-central F distributions with noncentrality parameters

$$\lambda_A(II) = \frac{[h_1(\mu_{11}-\mu_{21}) + h_2(\mu_{12}-\mu_{22})]^2}{2(h_1+h_2)(2\sigma^2)}$$

$$= \frac{[(h_1+h_2)(\alpha_1-\alpha_2)+h_1(\gamma_{11}-\gamma_{21})+h_2(\gamma_{12}-\gamma_{22})]^2}{2(h_1+h_2)(2\sigma^2)}$$

$$= [(h_1+h_2)\alpha_1 + (h_1-h_2)\gamma_{11}]^2/(h_1+h_2)\sigma^2 , \qquad (2.2.6)$$

and

$$\lambda_A(III) = \frac{[\mu_{11} - \mu_{21} + \mu_{12} - \mu_{22}]^2}{2(n_1^{-1} + n_2^{-1})(2\sigma^2)}$$

$$= \Delta_I^2/(n_1^{-1} + n_2^{-1})\sigma^2 \qquad (2.2.7)$$

Note that $\lambda_A(II)$ is a function of $\gamma_{11}$ as well as $\alpha_1$ and thus $\lambda_A(II)$ is not necessarily equal to zero when $H_0^A$ is true. Note also that $\lambda_A(III)$ does not depend on $\gamma_{11}$, so that $\lambda_A(III) = 0$ if $H_0^A$ is true.

The power of the F-tests is directly related to the noncentrality parameters. The power $1-\beta$ can be computed using equations given in Johnson and Katz (1970, p. 192).

Note that, in addition to the model parameters, the noncentrality parameters (and hence the powers) depend on the values of $n_1$ and $n_2$. Some specific aspects of the dependency are now discussed in the context of a fixed total sample size $n_{11} + n_{12} + n_{21} + n_{22} = 40$. Now $0 < n_1 + n_2 < 20$, so the possible points $(n_1, n_2)$ lie in the region as depicted in Figure 2.1. Selected points in Figure 2.1 are identified with the actual cell frequency distribution.

2.2.1 Bias Due to Interaction

It was pointed out earlier that the Type II test does not test $H_0^A: \bar{\mu}_{1.} = \bar{\mu}_{2.}$. Therefore, the rejection probability (type I error rate) will be greater than the prescribed $\alpha$, depending on the

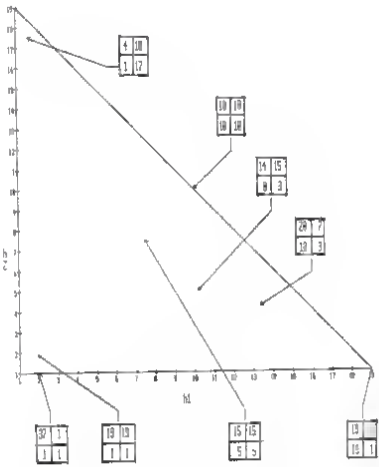



Figure 2 1 Harmonic Mean of $n_{11}$ and $n_{21}$ vs Harmonic Mean of $n_{12}$ and $n_{22}$. Region of possible values of $(h_1, h_2)$, with points identified corresponding to certain cell frequency patterns.

neg ilude if [illegible] the [illegible] error rate [illegible] a nominal size $\alpha = .05$ test with $\gamma_{11} = [illegible]$ is shown in Figure 2.2 as a function of $(h_1, h_2)$. Calculations for selected cell frequency patterns are shown in Table 2.1. Note that the true rejection rate can be as large as .2 for cell frequency distributions that are not uncommonly disperate.

2.2.2 Power of Tests

The bias in $F_A(II)$ noted above is due to the interaction parameter $\gamma_{11}$. If $\gamma_{11} = 0$, then both $F_A(II)$ and $F_A(III)$ are unbiased. Comparison of $\lambda_A(II)$ and $\lambda_A(III)$ where $\gamma_{11} = 0$ shows $\lambda_A(II) > \lambda_A(III)$ with equality only if $h_1 = h_2$. Powers of $F_A(II)$ and $F_A(III)$ for $\alpha_1 = .5\sigma$ are shown in Figures 2.3 and 2.4, respectively, as functions of $h_1$ and $h_2$ for .05 level tests. The inefficiency of $F_A(III)$ noted by Steel and Terrie (1960) and Bancroft (1968) is most pronounced when $h_1$ and $h_2$ differ greatly. Note that the power of $F_A(II)$ is constant for constant $h_1 + h_2$. See Table 2.1.

The presence of $\gamma_{11}$ in the noncentrality parameter for $F_A(II)$ effects the power as well as the type one error rate. For given $\alpha_1$, the rejection probability can (theoretically) range anywhere between the nominal $\alpha$ to 1 depending on the value of $\gamma_{11}$. The minimum rejection probability (= nominal $\alpha$) occurs when $(h_2 - h_1)\gamma_{11} = (h_1 + h_2)\alpha_1$. Large rejection probabilities occur for large positive values of $(h_1 - h_2)\gamma_{11}$. Figure 2.5 shows the dependence of the rejection probability on $(h_1, h_2)$ for $\alpha_1 = 1.0\sigma$ and $\gamma_{11} = .5\sigma$, for .05 level tests.

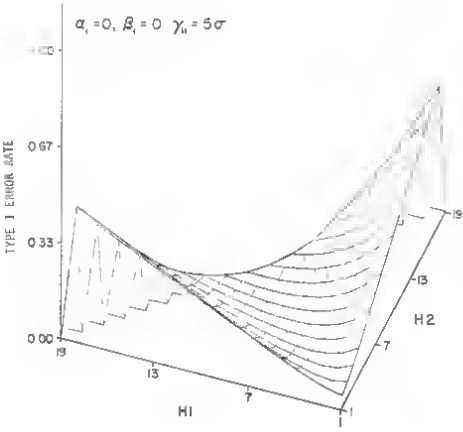



Figure 2 2 Type 1 Error Rate of $F_A(II)$. Rejection probability (bias) of $F_A(II)$

Table 2 1 Noncentrality Parameters ($\lambda$) and Selection Probabilities (Prob.) for Various Tests and Selected Sample Sizes. All tests are at the .05 level of significance, except the interaction test which is at the .25 level

| | | | | | | $Y_{11} = 5\sigma$ | | $\sigma = 5\sigma$ | | $\sigma = 5\sigma$ | | $\sigma = 1\sigma$, $Y_{11} = 5\sigma$ | | $Y = 5\sigma$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Type II Bias | | Type II Power | | Type III Power | | Interaction Effect on Type II Power | | Power of Interaction | |
| $n_{11}$ | $n_{21}$ | $h_1$ | $n_{12}$ | $n_{22}$ | $h_2$ | $\lambda$ | Prob | $\lambda$ | Prob | $\lambda$ | Prob | $\lambda$ | Prob | $\lambda$ | Prob. |
| 4 | 1 | 1.60 | 18 | 17 | 17.49 | 1.66 | .42 | 2.39 | .57 | 0.74 | .22 | 3.26 | .70 | 0.74 | .53 |
| 10 | 10 | 10.00 | 10 | 10 | 10.00 | 0.00 | .05 | 2.50 | .59 | 2.50 | .59 | 10.00 | .99 | 2.50 | [illegible] |
| 14 | 8 | 10.18 | 15 | 3 | 5.00 | 0.22 | .10 | 1.90 | .47 | 1.68 | .43 | 10.31 | .99 | 1.68 | [illegible] |
| 20 | 10 | 13.33 | 7 | 3 | 4.20 | 0.60 | .19 | 2.19 | .53 | 1.60 | .41 | 13.83 | .99 | 1.60 | .41 |
| 19 | 19 | 19.00 | 1 | 1 | 1.00 | 2.03 | .50 | 2.50 | .59 | 0.48 | .16 | 21.03 | .99 | 0.48 | .41 |
| 15 | 5 | 7.50 | 15 | 5 | 7.50 | 0.00 | .05 | 1.88 | .47 | 1.88 | .47 | 7.50 | .96 | 1.88 | .78 |
| 19 | 1 | 1.90 | 1 | 19 | 1.90 | 0.00 | .05 | 0.48 | .16 | 0.48 | .16 | 1.90 | .47 | 0.48 | .44 |
| 37 | 1 | 1.95 | 1 | 1 | 1.00 | 0.04 | .06 | 0.37 | .13 | 0.33 | .12 | 1.99 | .49 | 0.33 | [illegible] |

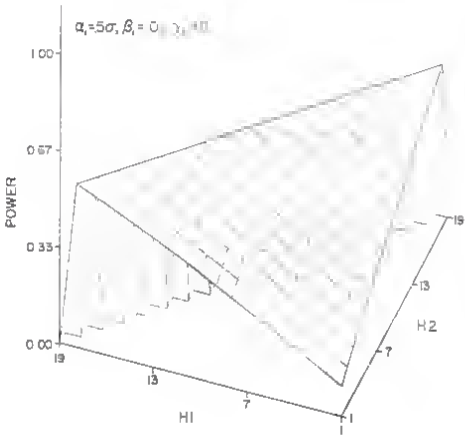



FIGURE 2.3 Power of $F_A(II)$ – Rejection probability of $F_A(II)$ –

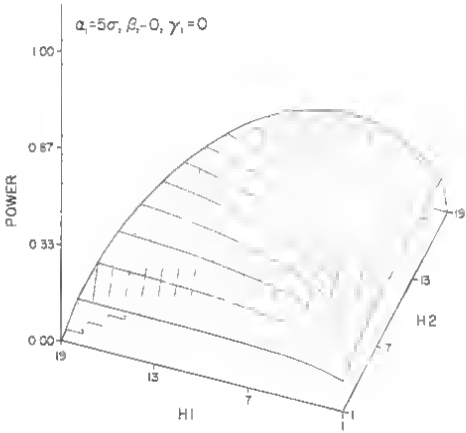



Figure 2.4 Power of $F_A$(III). Rejection probability of $F_A$(III).

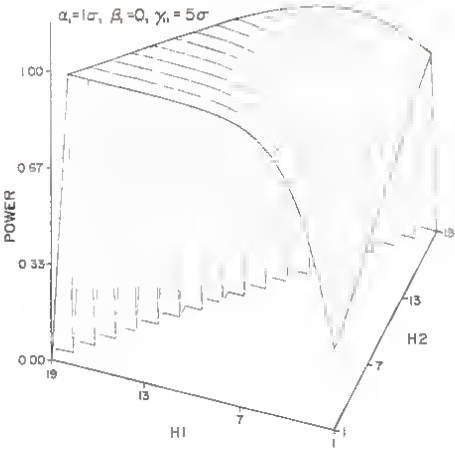



Figure 2.5 Rejection Probability of $F_A(II)$

As noted, Bancroft [illegible] suggests testing $H_0^{AB}$ $\alpha_1$ [illegible] comparing models (1) and (2) in (2.2.1)) at $\alpha = .25$ and recommends

1. testing $H_0^A$ with $F_A(II)$ if the test of $H_0^{AB}$ is nonsignificant; or
2. testing $H_0^A$ with $F_A(III)$ if the test of $H_0^{AB}$ is significant

provided the main effect hypothesis $H_0^A$ is still of interest. This recommendation attempts to utilize preliminary information about interaction to aid in the selection of the main effect test statistic. Figure 2.6 shows the power of the $\alpha = .25$ level interaction test with $\gamma_{11} = .50$ as a function of $h_1$ and $h_2$. Selected power computations are given in Table 2.1. The power of the interaction test has essentially the same form as the power of $F_A(III)$, and is thus most adversely affected when $h_1$ and $h_2$ differ greatly. Note that for the case $n_{11} = 20$, $n_{21} = 10$, $n_{12} = 7$, and $n_{22} = 3$, the power is only $1-\beta(\lambda) = .73$. Thus, for this particular design, even using Bancroft's suggested procedure, the probability is $\beta(\lambda) = .27$ of incorrectly assuming no interaction and thus being led to the use of $F_A(II)$. And this in turn could lead to excessive rejection probability if $H_0^A$ $\alpha_1 = 0$ is true.

2.2.3 The 2×b Case

The results for the 2×2 case extend directly to the 2×b case if one recognizes that $SS_A(II)$ and $SS_A(III)$ derive from the contrasts

$$C_A(II) = \Sigma^b h_j(\bar{y}_{1j.} - \bar{y}_{2j.})$$

and

$$C_A(III) = \Sigma^b (\bar{y}_{1j.} - \bar{y}_{2j.}) .$$

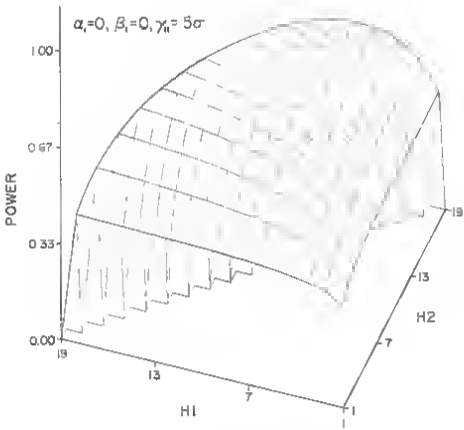



Figure 2.6 Power of Test for Interaction Rejection rate for $\alpha$ = .25 test of interaction

where $h_j = 2(n_{1j}^{-1} + n_{2j}^{-1})^{-1}$ is the harmonic mean of $n_{1j}$ and $n_{2j}$, $j=1,\ldots,b$. The corresponding noncentrality parameters are

$$\lambda_A(II) = 2[\sigma_1 \sum_{j=1}^{b} h_j + \sum_{j=1}^{b} h_j \gamma_{1j}]^2/(\sum_{j=1}^{b} h_j(2\sigma^2)) \qquad (2.2.8)$$

and

$$\lambda_A(III) = 2b^2 \sigma_1^2/(\sum_{j=1}^{b} h_j^{-1}(2\sigma^2)). \qquad (2.2.9)$$

Inspection of $\lambda_A(II)$ and $\lambda_A(III)$ reveals that the conclusion drawn for the $b=2$ case apply, as general phenomena, to the $b > 2$ case as well. That is, highly disparate $h_j$, resulting from small cell frequencies for some levels of $b$, create large bias in $F_A(II)$ and inefficiency in $F_A(III)$.

## 2.3 Conclusion

Unbalanced cell frequencies in a 2×2 classification affect the power of various tests in the analysis of variance. Structures for which the harmonic means $h_1$ and $h_2$ differ greatly are particularly disruptive. This condition typically occurs when there is a small cell frequency in one of the columns (levels of B) but not both, and results in bias in $F_A(II)$, and inefficiency in $F_A(III)$ and in the test for interaction. There does not appear to be an easy solution to the problem. However, the point should be made that a dogmatic rule such as "always use type II" or "always use type III" is not without pitfalls. Based on the limited calculations made here, it appears that bias in $F_A(II)$ may be a more severe problem than inefficiency in $F_A(III)$.

# CHAPTER 3
# ESTIMATORS IN LINEAR

## 3.1 Introduction

In this chapter, several biased estimators of the parameter vector in linear regression are discussed. A criterion for goodness must be established to allow comparison of the quality of one estimator versus another. Therefore in section 3.2, risk functions are introduced as a basis for comparison. In section 3.3 the widely used preliminary test estimator is discussed. In section 3.4 the groundwork is laid for "shrinking" an estimator toward a subspace. Two types of shrinkage estimators are then generalized to shrink toward a subspace.

## 3.2 Criterion for Goodness in Estimators

There are many attributes of an estimator (e.g. unbiasedness, efficiency, consistency, etc.), but there is rarely a single "best" estimator in terms of all of these qualities. Because of this, one must decide which attribute or combination of attributes is important, and choose an estimator accordingly.

One way to evaluate the goodness of an estimator is to study the "risk" associated with the use of the estimator. If the (K×1) parameter vector $\underline{\theta}$ is being estimated by the estimator $\hat{\underline{\theta}}$, then the risk associated with $\hat{\underline{\theta}}$ is the value of the risk function, $R(\hat{\underline{\theta}}, \underline{\theta})$. The risk function is defined by the statistician to reflect the importance of those attributes most preferred by the statistician.

For example, the statistician that favors estimators having small expected squared distance from the true estimator would use as the risk function:

$$R_1(\hat{\underline{\theta}}, \underline{\theta}) = E((\hat{\underline{\theta}} - \underline{\theta})'(\hat{\underline{\theta}} - \underline{\theta})). \qquad (3.2.1)$$

The statistician then chooses an estimator that has a small risk over all possible values of $\underline{\theta}$.

A more general risk function of which $R_1(\hat{\underline{\theta}}, \underline{\theta})$ is a special case, is:

$$R(\hat{\underline{\theta}}, \underline{\theta}) = E((\hat{\underline{\theta}} - \underline{\theta})'D(\hat{\underline{\theta}} - \underline{\theta}) \qquad (3.2.2)$$

where D is a symmetric (K×K) matrix of weighting constants. Hence the risk described in (3.2.2) is referred to as weighted risk when $D = I$. For $D = I$, $R(\hat{\underline{\theta}}, \underline{\theta})$ is $R_1(\hat{\underline{\theta}}, \underline{\theta})$.

In the factorial experiment, it may be that estimates of the cell means are to be used to get estimates of the marginal means, $\underline{\mu}_m$, as described in (1.1.9). In that case, a risk function measuring expected squared distance of the estimated marginal means from the true marginal means is needed. Such a risk function is

$$R_2(\underline{\theta}, \underline{\mu}) = E((\underline{\theta} - \underline{\mu})' D^{*} (\underline{\theta} - \underline{\mu})), \qquad (3.2.3)$$

where $\underline{\mu}$ is the vector of cell means as in (1.1.3), $\underline{\theta}$ is some estimator of $\underline{\mu}$, and

$$\underline{\mu}_M = \begin{pmatrix} b^{-1}(I_a \otimes \underline{1}'_b) \\ a^{-1}(\underline{1}'_a \otimes I_b) \end{pmatrix} \underline{\mu} = \begin{pmatrix} b^{-1}(I_a \otimes \underline{1}'_b) \\ a^{-1}(\underline{1}'_a \otimes I_b) \end{pmatrix} \underline{\mu} .$$

To see that this is appropriate, consider the unweighted risk of an estimator of marginal means ($\underline{e}_M$) that results from a particular estimator of cell means ($\underline{e}$) :

$$R_1(\underline{e}_M , \underline{\mu}_M) = E( (\underline{e}_M - \underline{\mu}_M)'(\underline{e}_M - \underline{\mu}_M) ) .$$

Using (11.8), $\underline{e}_M - \underline{\mu}_M = \begin{pmatrix} b^{-1}(I_a \otimes \underline{1}'_b) \\ a^{-1}(\underline{1}'_a \otimes I_b) \end{pmatrix} (\underline{e} - \underline{\mu})$ ,

it can be seen that $R_1(\underline{e}_M , \underline{\mu}_M) = R_2(\underline{e} , \underline{\mu})$ .

Using a risk function $R(\tilde{\underline{\theta}} , \underline{\theta})$ as a goodness criterion, $\hat{\underline{\theta}}$ is "better" than $\tilde{\underline{\theta}}$ if

1. $R(\hat{\underline{\theta}} , \underline{\theta}) \le R(\tilde{\underline{\theta}} , \underline{\theta})$ for all possible values of $\underline{\theta}$.

2. $R(\hat{\underline{\theta}} , \underline{\theta}) < R(\tilde{\underline{\theta}} , \underline{\theta})$ for at least one value of $\underline{\theta}$

It follows that the "best" estimator, if one exists, is better than all

other possible estimators. If the supremum of $R(\tilde{\underline{\theta}}, \underline{\theta})$ over all $\underline{\theta}$ is less than or equal to the supremum of $R(\bar{\underline{\theta}}, \underline{\theta})$ over all $\underline{\theta}$ for any $\bar{\underline{\theta}}$, then $\tilde{\underline{\theta}}$ is a minimax estimator.

## 3-3 Preliminary Test Estimators

Simply put, a preliminary test estimator is an estimator that takes one of two different forms depending on the outcome of a test of hypothesis.

If $\underline{x}$ is the vector of data, and the parameter vector $\underline{\theta}$ is being estimated, then a preliminary test estimator would take on the form:

$$\tilde{\underline{\theta}} = \begin{cases} \hat{\underline{\theta}}_1(\underline{x}) & \text{if } H_0 \text{ is rejected;} \\ \hat{\underline{\theta}}_2(\underline{x}) & \text{if } H_0 \text{ is accepted.} \end{cases} \qquad (3\text{-}3.1)$$

where $\hat{\underline{\theta}}_1(\underline{x})$ is an estimator of $\underline{\theta}$ that is appropriate in the case that $H_0$ is false, and $\hat{\underline{\theta}}_2(\underline{x})$ is an appropriate estimator of $\underline{\theta}$ when $H_0$ is true.

For general linear regression models the vector of unknown parameters $\underline{\beta}$ is often estimated using a preliminary test estimator. Suppose X is (N×K) rank K, $\underline{y}$ is (N×1), and $\underline{\beta}$ is (K×1) unknown. For the linear model

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}, \quad \underline{\epsilon} \sim N(\underline{0}, \sigma^2 I_N), \qquad (3\text{-}3.2)$$

the ordinary least squares estimate of $\underline{\beta}$ is $\underline{b} = (X'X)^{-1} X'\underline{y}$. It may be

suspected that $R\underline{\beta} = \underline{r}$, where R is (J×K) of rank $J \leq K$, and $\underline{r}$ is [illegible]. Recall from the section on parameter estimation in chapter 1, if $R\underline{\beta} = \underline{r}$ were known to be true, the restricted least squares estimator is

$$\bar{\underline{\beta}} = (I-MR)\underline{b} + M\underline{r} \ , \qquad (3.3.3)$$

where $M = S^{-1}R'(RS^{-1}R')^{-1}$ (This is the same form used by Bock, Yancey, and Judge (1973).) When a test of the hypothesis $H_0$: $R\underline{\beta} = \underline{r}$ is used to determine if $\underline{b}$ or $\bar{\underline{\beta}}$ will be used to estimate $\underline{\beta}$, a preliminary test estimator is being used:

$$\hat{\hat{\underline{\beta}}} = \begin{cases} \underline{b} & \text{if } H_0 \text{ is rejected.} \\ \bar{\underline{\beta}} & \text{if } H_0 \text{ is accepted} \end{cases} \qquad (3.3.4)$$

Let F denote the F statistic used to test $H_0$, and c denote the critical value used in the test. Using the standard indicator function notation, the preliminary test estimator is given in Bock, Yancey and Judge (1973) to be

$$\hat{\hat{\underline{\beta}}} = \underline{b} - I_{(0,c)}(F)\, M(R\underline{b} - \underline{r}) \qquad (3.3.5)$$

This may be rewritten as

$$\hat{\hat{\underline{\beta}}} = \bar{\underline{\beta}} + (1 - I_{(0,c)}(F))\, M(R\underline{b}-\underline{r}) \qquad (3.3.6)$$

A bibliography of studies on preliminary test estimators is given by Bancroft and Han (1977). Bock, Yancey and Judge (1973) describe the

properties of preliminary test estimators in the general linear model. Solove, Morris, and Radhakrishan (1972) criticize preliminary test estimators, calling them non-optimal. More recently, Bickal (1984) discusses the robustness of preliminary test estimators as compared to other biased estimators.

Several of these authors conclude that a preliminary test estimator is an undesirable choice of estimator. Nevertheless, preliminary test estimators are probably the most commonly used estimators for linear modeling.

## 3.4 Shrinkage Estimators

A shrinkage estimator of $\underline{\theta}$ is a functional combination of two estimators of $\underline{\theta}$, $\hat{\underline{\theta}}_1$ and $\tilde{\underline{\theta}}_2$ say. Typically, $\tilde{\underline{\theta}}_1$ is the maximum likelihood estimator, and $\hat{\underline{\theta}}_2$ is some other estimator. The estimator, $\hat{\underline{\theta}}_1$ might be unbiased, while $\tilde{\underline{\theta}}_2$ by itself might have a small variance, but be biased and inconsistent. By using the shrinkage estimator, a compromise between the good and bad properties of $\hat{\underline{\theta}}_1$ and $\tilde{\underline{\theta}}_2$ is attained.

The prototype of shrinkage estimators was proposed by James and Stein (1961). Variations of the James-Stein estimator have been proposed by Berger (1976a), Strawderman (1971), and Efron and Morris (1973), as well as many others. These estimators were originally proposed as compromises between $\underline{x}$, the ordinary least squares (OLS) estimator of $\underline{\theta}$, and some arbitrary fixed point $\underline{\theta}^0$, say. Efron and Morris (1973) write the general shrinkage estimator, componentwise as

$$\hat{\theta}_i = (1 - B_i)\, x_i + B_i\, \theta_i^0 , \quad i=1,2,\ldots,k , \qquad (3.4.1)$$

where k is the dimensionality of $\underline{\beta}$. The greatest advantages of using a shrinkage estimator such as this are realized when $\underline{c}$ is near $\underline{c}^0$. Thus the choice of $\underline{c}^0$ is quite a critical step in making good use of a shrinkage estimator in practice.

In a rejoinder to comments on his article, Stein (1962) suggested that a subset of the components of $\underline{\beta}$ could be estimated using a shrinkage estimator, with the balance of the components being estimated using the OLS estimator. Oman (1982) proposed that the OLS estimator could be shrunk toward a subspace rather than a fixed point. Lee (1983) discusses shrinking the OLS estimator of $\underline{\beta}$ toward a subspace in the regression context, where $\underline{\beta}$ is the parameter vector in the linear regression model. Judge and Bock (1978) give the form of the James Stein estimator that shrinks the OLS estimator of $\hat{\underline{\beta}}$ toward the subspace $\{\underline{\beta}: R\underline{\beta}=\underline{r}\}$, where $\underline{\beta}$ is the parameter vector in the linear regression model, and R is a J×K restriction matrix as introduced in section 1.1.

In this chapter, shrinkage estimators are considered that shrink the OLS estimator of $\underline{\beta}$ toward $\{\underline{\beta}: R\underline{\beta}=\underline{0}\}$ and take the form

$$\hat{\underline{\beta}} = (I - MR)\underline{b} + M\hat{\underline{\alpha}} , \qquad (3.4.2)$$

where $M=(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}$. In the above estimator, $\hat{\underline{\alpha}}$ is a shrinkage estimator of $R\underline{\beta}$, compromising between $R\underline{b}$ and $\underline{0}$, and $\underline{b}$ is the OLS estimator of $\underline{\beta}$ in a linear regression model.

The motivation for dealing with estimators in the form stated in (3.4.2) comes from the preliminary test estimator as stated in (3.3.6). One can see that the preliminary test estimator has the form

stated in (3.4.1). In the case of the preliminary test estimator, $\hat{\underline{\theta}}$ can take on one of two values: $\underline{0}$ or $R\underline{b}$. The shrinkage estimators of the form (3.4.2) that will be subsequently studied, compromise between the two extremes.

The following lemma will allow determination of sufficient conditions for minimaxity of the James-Stein estimators. Also it will allow generalization of the results obtained in the Monte Carlo experiments of chapter 4.

Lemma 3.4.1. For $\hat{\underline{\theta}}$ of the form given in (3.4.2), if $\hat{\underline{\theta}}$ is a function of $\underline{y}$ only through $R\underline{b}$ and MSE, then

$$R(\hat{\underline{\theta}}, \underline{\theta}) = E((I-MR)\underline{b}, (I-MR)\underline{\theta}) + E(M\tilde{\underline{\theta}}, MR\underline{\theta}), \qquad (3.4.3)$$

where $R(\hat{\underline{\theta}}, \underline{\theta}) = E((\tilde{\underline{\theta}}-\underline{\theta})' D (\hat{\underline{\theta}}-\underline{\theta}))$, as in (3.2.2). Recall that D is a symmetric (K×K) matrix of weighting constants.

Proof:

$$\tilde{\underline{\theta}} - \underline{\theta} = (I-MR)(\underline{b} - \underline{\theta}) + M(\hat{\underline{\sigma}} - R\underline{\theta}).$$

$$R(\hat{\underline{\theta}}, \underline{\theta}) = E((I-MR)\underline{b}, (I-MR)\underline{\theta}) + R(M\tilde{\underline{\sigma}}, MR\underline{\theta})$$
$$+ 2E((\hat{\underline{\sigma}}-R\underline{\theta})' M' D (I-MR)(\underline{b}-\underline{\theta})).$$

$$\mathrm{Cov}\left(\begin{pmatrix} R\underline{b} \\ (I - XS^{-1}X')\underline{y} \end{pmatrix}, (I - MR)\underline{b}\right) =$$

$$\begin{pmatrix} RS^{-1}X' \\ I - XS^{-1}X' \end{pmatrix} \Sigma (I - MR) S^{-1} /$$

$$\begin{pmatrix} RS^{-1}X'XS^{-1}(I - MR)' \\ (I - XS^{-1}X') XS^{-1} (I - MR)' \end{pmatrix} \sigma^2 =$$

Notice that $RS^{-1}X'XS^{-1}(I-MR)' = RS^{-1}-RS^{-1}(R'(RS^{-1}R')^{-1}RS^{-1}) =$ and $(I-XS^{-1}X')XS^{-1}(I-MR)' = (XS^{-1}-XS^{-1}X'XS^{-1})(I-MR)' = 0$ Therefo

$$\text{Cov}\left(\begin{pmatrix} R\underline{b} \\ (I - XS^{-1}X')\,\underline{y} \end{pmatrix}, (I - MR)\,\underline{b}\right) = 0$$

since $\begin{pmatrix} R\underline{b} \\ (I - XS^{-1}X')\,\underline{y} \end{pmatrix}$ and $(I - MR)\underline{b}$ are normally distributed

vectors, zero covariance implies independence. Since MSE is proportional to the norm of $(I - XS^{-1}X')\,\underline{y}$, $R\underline{b}$ and MSE are simultaneously independent of $(I - MR)\,\underline{b}$. If $\hat{\underline{\theta}}$ is a function of $\underline{y}$ only through $R\underline{b}$ and MSE, then $\hat{\underline{\theta}}$ is independent of $(I-MR)\underline{b}$. Using this independence,

$$E\left( (\hat{\underline{\theta}}-M\underline{\beta})'M' D (I-MR)\underline{b}) \right) = E\left( (\hat{\underline{\theta}}-M\underline{\beta})'M') D E\left( (I-MR)(\underline{b}-\underline{\beta}) \right) = 0.$$

Therefore, the lemma holds.

3.4.1 James-Stein Estimators

The James-Stein estimator for the mean of the multivariate $N_p(\underline{\theta}, \sigma^2 I)$ distribution ($p \geq 3$) was described by James and Stein (1961). Subsequent improvements were made in this estimator by

Baranchik (1964) and Stein (1966), who developed the positive part version of the James-Stein estimator. The James-Stein estimator and the James-Stein positive part estimator are now described for the case that $\sigma^2$ is unknown.

Suppose $\underline{x}_i \sim N_p(\underline{\theta}, \sigma^2 I)$, independently, $i=1,2,\ldots,N$. Also, assume $p \geq 3$ and $\sigma^2$ unknown. The MLE of $\underline{\theta}$ is

$$\bar{\underline{x}} = \left(\sum_{i=1}^{N} \underline{x}_i\right)/N \tag{3.4.4}$$

The James-Stein estimator as presented by Judge and Bock (1978, p. 179), may be written

$$\hat{\underline{\theta}}^S = (1-c/F)\,\bar{\underline{x}}\,, \tag{3.4.5}$$

where $c=(N-p)(p-2)/((N-p+2)p)$ and $F=\bar{\underline{x}}'\bar{\underline{x}}/p\hat{\sigma}^2$ is the likelihood ratio test statistic of $H_0: \underline{\theta} = \underline{0}$. The test statistic F has the noncentral F distribution with p and (N-p) degrees of freedom and noncentrality parameter $\lambda = \underline{\theta}'\underline{\theta}/\sigma^2$. The choice of c is made in such a way that $R_1(\hat{\underline{\theta}}^S, \underline{\theta})$ will be minimized. The discussion of the role of the constant c will be taken up shortly.

The James-Stein estimator as given in (3.4.5) may be restated as

$$\hat{\underline{\theta}}^S = (1 - c/F)\,\bar{\underline{x}} + (c/F)\,\underline{0}\,, \tag{3.4.6}$$

or written componentwise as

$$\hat{\theta}_i^S = (1 - c/F)\,\bar{x}_i + (c/F)\,0\,, \quad i=1,2,\ldots,N \tag{3.4.7}$$

[illegible] that [illegible] tr[illegible] [illegible] Efron and Morris (1973) as stated in (3.4.1)

The constant c, called the shrinkage constant, controls the degree of influence that the data, through the F statistic, will be allowed to have on the value of the estimator. For example, if c were chosen to be zero, then even if the F test indicates that $\underline{\theta}$ is $\underline{0}$, the estimator $\hat{\underline{\theta}}^S$ will be $\bar{\underline{x}}$. On the other hand, if c were chosen to be 100, say, then depending on the observed value of the F statistic, $\hat{\underline{\theta}}^S$ would likely be quite different from $\bar{\underline{x}}$.

Depending upon the chosen criterion for goodness that is being used to evaluate estimators, reasonable choices of c can be determined. For example, if unbiasedness is the ultimate criterion for goodness, then c=0 would be the right choice of c. The traditional criterion for goodness is the risk function $R_1(\cdot,\cdot)$ as described in (3.2.1). Judge and Bock (1978, pp. 177-179), analytically demonstrate that the choice of c which leads to the best James-Stein estimator, when $R_1(\cdot,\cdot)$ is used as the risk function in evaluating goodness of the estimator, is $c = (N-p)(p-2)/(p(N-p+2))$.

If rather than a compromise between $\bar{\underline{x}}$ and $\underline{0}$, a compromise between $\bar{\underline{x}}$ and $\underline{\theta}_0$ is desired, the James-Stein estimator for shrinking $\bar{\underline{x}}$ toward $\underline{\theta}_0$ is written

$$\hat{\underline{\theta}}^S = (1-c/F)\,\bar{\underline{x}} + (c/F)\,\underline{\theta}_0 , \qquad (3.4.8)$$

where $c=(N-p)(p-2)/((N-p+2)p)$, as in (3.4.5). The F statistic is now the

one that tests $H_0: \underline{\theta} = \underline{0}$, rather than $H_0: \underline{\theta} = \underline{\theta}$. [illegible] a extra[illegible] discussed in Judge and Bock (1978, p. 180).

In terms of $R_1(\cdot,\cdot)$, the expected squared distance risk function, the James-Stein estimator is better than the MLE. Judge and Bock (1978, pp. 85 and 175) report the risk function of the James-Stein estimator to be

$$R_1(\hat{\underline{\theta}}^S, \underline{\theta}) = \sigma^2\left(p - \frac{(p-2)^2(N-p)}{(N-p+2)}\right) E(1/G) , \qquad (3.4.9)$$

where $G \sim \chi^2(p,\lambda)$. It is well known that the risk of the MLE (which is the OLS estimator in this case) is

$$R_1(\bar{\underline{x}}, \underline{\theta}) = \sigma^2 p , \qquad (3.4.10)$$

Judge and Bock (1978) give the graph of these risk functions for increasing $\lambda$ as shown in Figure 3.1. Notice that $R_1(\hat{\underline{\theta}}^S, \underline{\theta}) < R_1(\bar{\underline{x}}, \underline{\theta})$, for finite $\lambda$ and $p \geq 3$. Also, $\hat{\underline{\theta}}^S$ makes its biggest gain over $\bar{\underline{x}}$ when $\underline{\theta} = \underline{0}$.

One unfortunate property of the estimator in (3.4.5) is that $c/F$ is not restricted to be less than or equal to 1. There is a nonzero probability that $F<c$, allowing $c/F>1$. Thus $\hat{\underline{\theta}}^S$ can fall outside the interval $[\underline{0},\bar{\underline{x}}]$. To alleviate this possibility, the James-Stein positive part estimator was developed by Baranchik (1964) and Stein (1966).

The James-Stein positive part estimator as reported by Judge and Bock (1978, pp. 182-184) may be written

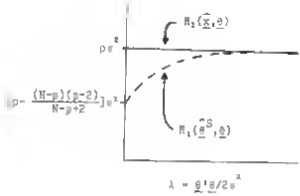



Figure 3.1 Risk of OLS and James-Stein Estimators. (Adapt[illegible] and Book, 1978, p. 178.)

$$\hat{\underline{\theta}}^{S+} = [\ \ \ \

[illegible]st th[illegible]antages of [illegible] are greatest [illegible] large.

Judge and Bock (1978, p. 24) [illegible] the James-Stein and James-Stein positive part estimators to allow shrinking $\underline{b}$ toward $\{\underline{R}\underline{\beta}=\underline{r}\}$. In this dissertation, the eventual intent is to shrink interaction in the factorial experiment toward zero. The constraint of additivity (no interaction) may be written using the appropriate restriction matrix as $\underline{R}\underline{\mu}=\underline{0}$. Therefore, the James-Stein estimators discussed subsequently will shrink $\underline{b}$ toward $\{\underline{R}\underline{\beta} = \underline{0}\}$.

Suppose $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, where $\underline{y}$ is $(N\times 1)$, $X$ is $(N\times K)$ of rank $K$, $\underline{\beta}$ is the $(K\times 1)$ vector of parameters, and $\underline{\varepsilon} \sim N_N(\underline{0}, \sigma^2 I)$. The OLS estimator of $\underline{\beta}$ is $\underline{b}=S^{-1}X'\underline{y}$. Assuming that $R\underline{\beta} = \underline{0}$, where $R$ is $(J\times K)$ of rank $J$, the restricted estimator of $\underline{\beta}$ is $\bar{\underline{\beta}} = \underline{b} - MR\underline{b}$, where $M = S^{-1}R'(RS^{-1}R')^{-1}$. To insure that $\underline{b}$ and $\bar{\underline{\beta}}$ can be computed it is required that $J\leq K\leq N$. Rather than take $\underline{b}$ or $\bar{\underline{\beta}}$ as the estimator, the generalized James-Stein estimator is a compromise between $\underline{b}$ and $\bar{\underline{\beta}}$. Since this generalized James-Stein estimator is essentially shrinking $R\underline{b}$ toward $\underline{0}$, it is required that $J$, the rank of $R\underline{b}$, be at least 3. The generalized James-Stein estimator as given by Judge and Bock (1978) may be written

$$\hat{\underline{\beta}}^S = (1-c/F)\,\underline{b} + (c/F)\,\bar{\underline{\beta}}, \qquad (3\text{-}4\text{-}12)$$

where $c$ is a shrinkage constant, $3\leq J\leq K\leq N$, and

$$F = \frac{(R\underline{b}-\underline{0})'(R(X'X)^{-1}R')^{-1}(R\underline{b}-\underline{0})(N-K)}{J(\underline{y}-X\underline{b})'(\underline{y}-X\underline{b})}$$ is the likelihood ratio test statistic of $H_0$: $R\underline{b} = \underline{0}$. This test statistic has the F distribution with J and (N-K) degrees of freedom and noncentrality parameter

$$\lambda = (R\underline{\beta}-\underline{0})'(R(X'X)^{-1}R')^{-1}(R\underline{\beta}-\underline{0})/2\sigma^2 \qquad (3.4.13)$$

The estimator given in (3.4.12) is written in a form that is concise and easy to see as being a compromise between $\underline{b}$ and $\hat{\underline{\beta}}$. This estimator may be rephrased as

$$\hat{\underline{\beta}}^S = (I-MR)\underline{b} + M\hat{\underline{\alpha}}^S \qquad (3.4.14)$$

where $\hat{\underline{\alpha}}^S = (1-c/F)\,R\underline{b}$. The equivalency can be seen by observing the following:

$$\begin{aligned}
\hat{\underline{\beta}}^S &= (1-c/F)\underline{b} + (c/F)(I-MR)\underline{b} \qquad \text{(from (3.4.12))},\\
&= [(1-c/F)(I-MR)\underline{b} + (1-c/F)MR\underline{b}] + [(c/F)(I-MR)\underline{b} + (c/F)\underline{0}]\ ,\\
&= (I-MR)\underline{b} + M[(1-c/F)R\underline{b}]\ ,\\
&= (I-MR)\underline{b} + M\,\hat{\underline{\alpha}}^S.
\end{aligned}$$

Notice that $\hat{\underline{\alpha}}^S$ is the James-Stein estimator of $R\underline{\beta}$ that compromises between $R\underline{b}$ and $\underline{0}$. Writing $\hat{\underline{\beta}}^S$ as in (3.4.14) makes it easy to see that $\hat{\underline{\beta}}^S$ is an estimator of the form described in (3.4.2). Throughout this work the James-Stein estimator will be phrased in either of these two ways.

Again, the shrinkage constant c determines how much the data, through the F statistic, will be allowed to influence the value of the estimator away from the OLS estimator. In this more complex setting, the optimal choice of value for c is not so clear. The constant c must be chosen to allow $\hat{\underline{\theta}}^S$ to be best with respect to the weighted risk function

$$R(\hat{\underline{\theta}}^S, R\underline{\theta}) = E((\hat{\underline{\theta}}^S - R\underline{\theta})'M'M(\hat{\underline{\theta}}^S - R\underline{\theta})) . \qquad (3.A.15)$$

To see that this is the case, note the form of $\hat{\underline{\theta}}^S$

$$\hat{\underline{\theta}}^S = (1 - c/F)R\underline{b} .$$

In an expanded form

$$\hat{\underline{\theta}}^S = R\underline{b} - c\,\frac{R\underline{b}}{(\underline{b}-\underline{\theta})'R'(RS^{-1}R')^{-1}R(\underline{b}-\underline{\theta})/(MSE\cdot J)} .$$

By inspection, $\hat{\underline{\theta}}^S$ is a function of $\underline{y}$ only through MSE and $R\underline{b}$. Therefore, Lemma 3.4.1 gives the result

$$R_1(\hat{\underline{\theta}}^S, \underline{\theta}) = R_1((I-MR)\underline{b}, (I-MR)\underline{\theta}) + R_1(M\hat{\underline{\theta}}^S, MR\underline{\theta}) .$$

In the special case that M'M is proportional to I then $c=(N-K)(J-2)/(J(N-K+2))$ allows the best James-Stein estimator of $\underline{\theta}$ (in terms of $R_1(\cdot,\cdot)$). In other cases this choice may not be best. At the end of this section, conditions on the choice of c are given which insure the James-Stein estimator will be minimax.

A corresponding generalized James-Stein positive part [illegible] implied but not explicitly stated by Judge and Bock (1973). The generalized James-Stein positive part estimator may be written

$$\tilde{\underline{g}}^{S+} = (1 - \min(1, c_1/F)\ )\ \underline{b} + \min(1, c_1/F)\ \hat{\underline{g}}\ , \qquad (3.4.16)$$

where $c_1$ is a shrinkage constant, and F is as described in (3.4.12). The estimator in (3.4.16) may also be written in the form of (3.4.2), as follows

$$\tilde{\underline{g}}^{S+} = (I-MR)\underline{b} + M\ \underline{a}^{S+}\ , \qquad (3.4.17)$$

where $\hat{\underline{a}}^{S+} = (1-\min(1,\frac{c_1}{F}))R\underline{b}$. The constant $c_1$ plays precisely the same role as the constant in (3.4.11). Any choice of $c_1$ in the interval

$$[\ (J-2)(N-K)/((N-K+2)J), 2(J-2)(N-K)/((N-K+2)J)\ ]\ ,$$

will lead to a minimax estimator of $\underline{g}$ if M'M is proportional to I. Otherwise, some values of $c_1$ in the above interval will not necessarily lead to a minimax estimator of $\underline{g}$. At the end of this section, conditions for minimaxity are given.

One drawback of these James-Stein estimators described in (3.4.12) and (3.4.16) is that every element in $R\underline{b}$ is adjusted toward 0 by the same amount. Intuition (and Empirical Bayes methodology) suggests that the variance of $R\underline{b}$ should be allowed to influence the amount of shrinkage for each element of $R\underline{b}$ toward 0. The Efron-Morris estimator

that is discussed in the next section shrinks each element of $R\underline{b}$ by an amount dependent on the variance of the element.

The estimator in (3.4.16) will be applied in examples for the factorial experiment in chapter 3. Also, $\hat{\underline{\beta}}^{S+}$ will be compared with other estimators in a simulation.

<u>Conditions sufficient for minimaxity of the James-Stein estimator with respect to $R_1(\cdot,\cdot)$.</u> The James-Stein estimator that shrinks $\underline{b}$ toward the subspace $\{\underline{\beta}: R\underline{\beta}=\underline{0}\}$ as stated in (3.4.14) is

$$\hat{\underline{\beta}}^{S} = (I-MR)\underline{b} + M\hat{\underline{\alpha}}^{S} . \qquad (3.4.18)$$

As before, $\underline{\beta}$ is the parameter vector in a linear regression equation, and $\underline{b}$ is the OLS estimator of $\underline{\beta}$. Also as before, $R$ is a $J\times K$ restriction matrix of rank $J\leq K$, $S=X'X$, $M=S^{-1}R'(RS^{-1}R')^{-1}$ and

$$\hat{\underline{\alpha}}^{S} = \frac{c}{F}\underline{0} + (1-\frac{c}{F})R\underline{b} , \qquad (3.4.19)$$

where $F$ is the test statistic for $H_0: R\underline{\beta}=\underline{0}$, and $c$ is a constant. From Lemma 3.4.1 it can be seen that $\hat{\underline{\beta}}^{S}$ will be minimax with respect to $R_1(\cdot,\cdot)$ if $\hat{\underline{\alpha}}^{S}$ is minimax with respect to the weighted risk function

$$R(\hat{\underline{\alpha}}, R\underline{\beta}) = E[(\hat{\underline{\alpha}}-R\underline{\beta})'M'M(\hat{\underline{\alpha}}-R\underline{\beta})] . \qquad (3.4.20)$$

From Judge and Bock (1978, p. 242), it can be found that the conditions for $\hat{\underline{\alpha}}^{S}$ to achieve the desired minimaxity are

$$\| L \| \le \frac{2(\mathrm{tr}(MRS^{-1})d_L^{-1}-2)(N-K)}{(N-K+2)\ J} ,$$

and

$$\mathrm{tr}(MRS^{-1}) > 2d_L ,$$

where $d_L$ is the largest characteristic root of $MRS^{-1}$. The conditions stated in (3.4.21) are sufficient to insure that $\hat{\theta}^S$ is minimax. Because the James-Stein positive part estimator provides uniform improvement over the James-Stein estimator, these conditions (substituting $c_1$ for c) are also sufficient for the James-Stein positive part estimator to be minimax.

3.4.2 Empirical Bayes

To understand the empirical Bayes approach to parameter estimation, recall the classical Bayes parameter estimation technique.

1. A prior distribution of the parameter $\theta$ is proposed; $g(\theta)$.
2. Data are collected, and the likelihood function is used to summarize the information obtained from the data, x, about the parameter $\theta$. The likelihood function is the conditional distribution of x given $\theta$; $f(x|\theta)$.
3. Using the prior distribution and the likelihood function a posterior distribution of $\theta$ given x is computed; $h(\theta|x)$.

When the prior distribution of $\theta$ is $N(\theta_0, \gamma^2)$ and the likelihood function is the density function for a $N(\theta, \sigma^2)$ distribution, the posterior distribution is

$$N\left(\frac{x^{[illegible]}}{\gamma^{[illegible]}}, \frac{[illegible]}{[illegible]+\sigma^2}\right) \qquad [illegible]$$

All inferences about $\theta$ are drawn f[illegible] the posterio[illegible] distribut[illegible] Specifically, the posterior mean is used as a point estimator for $\theta$

The empirical Bayes approach allows the collected data to influence the formulation of the prior distribution If a prior distribution can be postulated up to an unknown parameter, then the unknowe parameter is estimated from the data collected. This estimation of parameters for the prior distribution takes some of the pressure off of the statistician, for the most difficult step in the classical Bayes estimation technique is the statement of a realistic prior distribution.

Efron and Morris (1973) introduced an empirical Bayes estimator for the k dimensional parameter vector $\underline{\theta}$, with $k\geq 3$. The prior distribution is taken to be $N(\underline{0},AI)$, where A is an unknown scalar. Observed data are used to estimate A, thus the data help determine the exact formulation of the prior distribution A realization of the random variable $\underline{x}\sim N(\underline{\theta},D)$ is observed, where D is a positive definite, diagonal covariance matrix, assumed known The elements of D are $D_1,D_2,\ldots,D_k$. If A was known, the posterior distribution of $\underline{\theta}$ given $\underline{x}$ would be, written elementwise,

$$\theta_i|x_i \sim N((1-B_i)x_i, (1-B_i)D_i), \qquad (3.4.23)$$

where $B_i=D_i/(A+D_i)$. The Bayes estimate of $B_i$ would be the posterior mean Since A is not known, the empirical Bayes strategy requires that an estimate of A be substituted into the Bayes estimate of $\theta_i$

In the estimator proposed by Efron and Morris (1973), the data vector $\underline{x}$ is used to obtain estimates for $D_i$, $i=1,2,\ldots,k$, which is a function of A as in (3.4.23). The estimator of $\underline{\theta}$ written elementwise is

$$\hat{\theta}_i^E = (1 - \hat{B}_i)x_i \ , \qquad (3.4.24)$$

where $\hat{B}_i$ is an estimate of $D_i/(A+D_i)$. Specifically,

$$\hat{B}_i = \frac{(\hat{d}_i^* - 4)}{\hat{d}_i^*} \ \frac{D_i}{\hat{A}_i + D_i} \ . \qquad (3.4.25)$$

In (3.4.25), $\hat{A}_i$ will be an estimate of A which differs for each $i=1,2,\ldots,k$. The ratio $(\hat{d}_i^* - 4)/\hat{d}_i^*$ is an adjustment factor, where

$$\hat{d}_i^* = 2(\hat{A}_i + D_i)^2 \sum_{j=1}^{k} d_j/2(\hat{A}_i + D_j)^2 \ , \qquad (3.4.26)$$

and $d_i=3$, and $d_j=1$, $j\neq i$. The values of $\hat{A}_i$ are obtained by iteratively solving the equation

$$\hat{A}_i = \frac{\sum_{j=1}^{k} E_j(d_j/2(\hat{A}_i + D_j)^2)}{\sum_{j=1}^{k}(d_j/2(\hat{A}_i+D_j)^2)} \ , \ i=1,2,\ldots,k, \qquad (3.4.27)$$

where $E_j=(x_j^2-d_jD_j)/d_j$. If $D_1,D_2,\ldots,D_k$ are also unknown, then estimates of $D_1,D_2,\ldots,D_k$ must be substituted into (3.4.25) through (3.4.27). Efron and Morris (1973) state that this class of estimators includes the James-Stein estimator as the special case that $D_1=D_2=\ldots=D_k$ (all known).

As an example of the situation for which this Efron-Morris estimator is designed, consider the model for unequal observations from

k populations

$$Y_{ij} = \theta_i + \varepsilon_{ij} \quad i=1,2,\ldots,k \text{ and } j=1,2,\ldots,n_i$$

where every $\varepsilon_{ij}$ is i.i.d. $N(0,\sigma^2)$, $\sigma^2$ unknown. Let the Bayesian prior for $\theta_i$ be $\theta_i$ are i.i.d $N(0,\sigma_A^2)$, $\sigma_A^2$ unknown. Then, consider

$$x_i = \left( \sum_{j=1}^{n_i} Y_{ij} \right)/n_i \ , \ i=1,2,\ldots,k \qquad (3.4.23)$$

Recall that $x_i$ is distributed $N(\theta_i, \sigma^2/n_i)$, independent. Referring back to (3.4.23), the prior distribution variance A is in this example $\sigma_A^2$. The diagonal elements of D are in this example $\sigma^2/n_i$, $i=1,2,\ldots,k$. Since $\sigma^2$ is unknown, it must be estimated from the data. Estimation of $\underline{\theta}$ proceeds using the formulas in (3.4.24) through (3.4.27).

As an alternative to estimating $\sigma_A^2$ iteratively, with a different estimate of $\sigma_A^2$ for each element in the parameter vector, a closed form estimator for $\theta_i$ is now proposed.

If both $\sigma^2$ and $\sigma_A^2$ were known, the Bayesian estimate of $\theta_i$ would be

$$\hat{\theta}_i = \left(1 - \frac{\sigma^2/n_i}{(\sigma_A^2 + \sigma^2/n_i)}\right) x_i, \ i=1,2,\ldots,k \ , \qquad (3.4.30)$$

which can be written,

$$\hat{\theta}_i = \left(1 - \frac{1}{(n_i\sigma_A^2/\sigma^2 + 1)}\right) x_i \ , \ i=1,2,\ldots,k \qquad (3.4.31)$$

Since $\sigma^2_A/\sigma^2$ is unknown, the data must [illegible]

unbiased estimator of $v = \sigma^2_A/\sigma^2$ is

$$\bar{v} = \left\{ \frac{(N-k-2)\left(\sum_{i=1}^{k} n_i x_i^2\right)/k}{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - x_i)^2\right)} - 1 \right\} k/N , \qquad (3.4.32)$$

where $N = \sum_{i=1}^{k} n_i$

To see that $\bar{v}$ is unbiased for $v$, recall

$$\left( \frac{\sum_{i=1}^{k} n_i x_i^2/k}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-x_i)^2} \right) \times \left( \frac{(N-k)\sigma^2}{\sigma^2 + N\sigma_A^2/k} \right) \qquad (3.4.33)$$

has the central F distribution, with k numerator degrees of freedom, and N-k denominator degrees of freedom. The expected value of a F distributed random variable, with N-k denominator degrees of freedom is (N-k)/(N-k-2). Applying algebra to the expected value of the expression in (3.4.33),

$$E\left( \frac{\sum_{i=1}^{k} n_i x_i^2/k}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-x_i)^2} \right) = \frac{\sigma^2 + N\sigma_A^2/k}{(N-k)\sigma^2} \times \frac{N-k}{N-k-2}$$

$$= \frac{(1 + N\sigma_A^2/k\sigma^2)}{(N-k-2)} . \qquad (3.4.34)$$

Solving for $\sigma_A^2/\sigma^2$,

$$E(\bar{v}) = \sigma_A^2/\sigma^2 \qquad (3.4.35)$$

Substituting $\hat{v}$ for $\sigma_A^2/\sigma^2$ [illegible]

$$\hat{\hat{\theta}}_i = (1 - \frac{\text{constant}}{1 + n_i \hat{v}}) \times [\text{illegible}] \qquad (3.4.36)$$

where constant=(N-k-2)(k-2)/(N-k+2)k. The constant is included to cause this estimator to agree with the James-Stein estimator when all the $n_i$'s are equal.

The value of using this modification is that it is noniterative, allowing easy computation in practice. Ghosh and Meeden (1986) use a closed form estimator of the variance ratio, v, to develop an empirical Bayes estimate of the mean vector in a finite population sampling problem. Their estimator differs from (3.4.31) only by a constant.

For both the Efron-Morris estimator and the modified estimator in (3.4.36), a positive part version can be developed. As with other shrinkage estimators, the positive part version would be recommended for application.

The empirical Bayes estimator in (3.4.36) can be generalized to compromise between $\underline{b}$ and $\bar{\underline{\theta}}$ following the estimator form in (3.4.2). As always, let $\underline{y} = X\underline{\theta} + \underline{\epsilon}$, where $\underline{y}$ is (N×1), X is (N×K) of rank K, $\underline{\theta}$ is the (K×1) vector of parameters, and $\underline{\epsilon} \sim N_N(\underline{0}, \sigma^2 I)$. Further, assume that the prior distribution of $R\underline{\theta}$ is $N_J(\underline{0}, \sigma_R^2 I)$, where R is (J×K) of rank J such that $RR' = I$. If $\sigma^2$ and $\sigma_R^2$ were known, the posterior distribution of $R\underline{\theta}$ given $R\underline{b}$ would be

$$R\underline{\theta} \mid R\underline{b} \sim N(\hat{\underline{\theta}}, V). \qquad (3.4.37)$$

where $\hat{\underline{a}} = R\underline{b} - ( \sigma_R [illegible] I)^{-1} R\underline{b}$, and

$V = (I-(I+(\sigma_R^2/\sigma^2)[RS^{-1}R\ ]^{-1})^{-1})RS^{-1}R\ \sigma^2$ To see that this is true, take $P$ to be an orthogonal transformation matrix that diagonalizes $Var(R\underline{b})$. That is

$$P(RS^{-1}R')P' = D = diag( D_1, D_2, \ldots, D_J). \qquad (3.4.38)$$

The prior distribution for $\underline{Y} = PR\underline{a}$ will be

$$\underline{Y} \sim N( \underline{0} , I\sigma_R^2) \qquad (3.4.39)$$

The distribution of $\underline{t} = PR\underline{b}$ given $\underline{Y}$ will be

$$\underline{t} \mid \underline{Y} \sim N( \underline{Y} , D\sigma^2 ) \qquad (3.4.40)$$

The posterior distribution of $\underline{Y} \mid \underline{t}$ is then

$$\underline{Y} \mid \underline{t} \sim N( \dot{\underline{Y}} , W ) , \qquad (3.4.41)$$

where the elements of $\underline{Y}$ are

$$Y_i = (1 - (1 + D_i^{-1}\sigma_R^2/\sigma^2)^{-1} ) t_i \qquad (3.4.42)$$

The posterior variance $W = diag ( W_1, W_2, \ldots, W_J )$ Each diagonal element in $W$ has the form

$$W_i = (1 - [illegible]) \qquad (3.4.43)$$

In matrix notation,

$$\underline{Y} = (I - [illegible]) \qquad (3.4.44)$$

and

$$W = (I-(I+D^{-1}\sigma_R^2/\sigma^2)^{-1})D_0 \qquad (3.4.45)$$

Transforming back, the posterior distribution of $R\underline{\beta}$ given $R\underline{b}$ is

$$R\underline{\beta} \mid R\underline{b} \sim N(\underline{\alpha}, V), \qquad (3.4.46)$$

where

$$\begin{aligned}\underline{\alpha} &= P'(I - (I+D^{-1}\sigma^2_R/\sigma^2)^{-1})PR\underline{b} \\ &= (P'P - (P^{-1}(I+D^{-1}\sigma^2_R/\sigma^2)(P')^{-1})^{-1})R\underline{b} \\ &= (I - (P'(I+D^{-1}\sigma^2_R/\sigma^2)P)^{-1})R\underline{b} \\ &= (I - (P'P + (P'DP)^{-1}\sigma^2_R/\sigma^2)^{-1})R\underline{b} \\ &= (I - (I + (RS^{-1}R')^{-1}\sigma^2_R/\sigma^2)^{-1})R\underline{b}, \qquad (3.4.47)\end{aligned}$$

and $V = P'WP = (I-(I + (RS^{-1}R')^{-1}\sigma^2_R/\sigma^2)^{-1}(RS^{-1}R')\sigma^2$, as in (3.4.37).

An estimator of $\underline{\beta}$ of the form given in (3.4.2), using $\underline{\alpha}$ as the estimator of $R\underline{\beta}$ is

$$\underline{\hat{\beta}} = (I - MR)\underline{b} + M\left(I - \left(\frac{\sigma^2_R}{\sigma^2}(RS^{-1}R')^{-1}+I\right)^{-1}\right)R\underline{b} \qquad (3.4.48)$$

Written another way,

$$\tilde{\beta} = \underline{b} - M\left(\frac{\sigma_R^2}{\sigma^2}(RS^{-1}R' + I)\right)^{-1}R\underline{b} \qquad (3.4.49)$$

where $\underline{b}$ is the OLS estimate of $\underline{\beta}$, $S^{-1} = (X'X)^{-1}$, and $M = S^{-1}[S^{-1} + R'R]^{-1}$ (as in other sections). When $\sigma^2$ and $\sigma_R^2$ are unknown, an empirical Bayes estimator can be developed by substituting an estimate of $\sigma_R^2/\sigma^2$ into (3.4.49). Using an unbiased, closed form estimate of $\sigma_R^2/\sigma^2$, an empirical Bayes estimator would be

$$\hat{\underline{\beta}}^E = \underline{b} - gM(\hat{v}(RS^{-1}R') + I)^{-1}R\underline{b}, \qquad (3.4.50)$$

where $\hat{v} = [(N-K-2)F/(N-K) - 1]\,J/tr((RS^{-1}R')^{-1})$ is an estimator of $\sigma_R^2/\sigma^2$, $g = (N-K-2)(J-2)/(N-K+2)J$, and

$F = [(R\underline{b})'(RS^{-1}R')^{-1}R\underline{b}(N-K)]/[J(\underline{y}-X\underline{b})'(\underline{y}-X\underline{b})]$ is the likelihood ratio test statistic of $H_0: \sigma_R^2 = 0$. The value of the constant g was chosen to allow $\hat{\underline{\beta}}^E$ to be the same as the James-Stein estimator when M'M is proportional to I.

The estimate of $\sigma_R^2/\sigma^2$ can be shown to be unbiased using the same type argument as was given in (3.4.33) through (3.4.35).

A positive part version of this estimator, constraining $\hat{\beta}_i^E$ to $[\bar{\beta}_i, b_i]$ would be

$$\hat{\beta}_i^{E+} = \begin{cases} \hat{\beta}_i^E & \text{if } G>0 \text{ and } H>0, \\ \bar{\beta}_i & \text{if } G\leq 0, \\ b_i & \text{if } H\leq 0, \end{cases} \qquad (3.4.51)$$

where $G = ($ [illegible] $)$, $H = ( \hat{\theta}_i^{B} - b$ [illegible]

$\tilde{\theta}_i$ is the estimate for $\theta_i$ assuming $\sigma^2_{\theta}=0$, for $i=1,2,\ldots,N$. The choice of $g$ in $\hat{\underline{\theta}}^{E+}$ to force $\hat{\underline{\theta}}^{E+}$ to correspond with $\hat{\underline{\theta}}^{S+}$ in (3.4.17) when $M'M$ is proportional to $I$ is $g=(N-K-2)c_1/(N-K)$. Recall that $c_1$ is the constant used in $\hat{\underline{\theta}}^{S+}$.

The three conditions listed in (3.4.51) are a tedious way of determining the status of each element in the empirical Bayes estimator with respect to the interval $[b_i,\tilde{\theta}_i]$. The first condition is true for $\hat{\theta}_i^{E}$ inside the interval. The second condition is true for $\hat{\theta}_i^{E}$ violating the $\tilde{\theta}_i$ end of the interval. The third condition is true for $\hat{\theta}_i^{E}$ violating the $b_i$ end of the interval.

The estimator in (3.4.51) will be applied in examples for the factorial experiment in chapter 3. Also, $\hat{\underline{\theta}}^{E+}$ will be compared to other estimators in a simulation.

# CHAPTER 4
# TWO-WAY FIXED EFFECTS FACTORIAL EXPERIMENT

In chapter 3, several shrinkage estimators were discussed in the generalized framework of linear regression. In this chapter the estimators described in chapter 3 are applied to the two-way fixed effects factorial experiment. This application parallels the work of Lee (1983), but is approached from a slightly more applied point of view. After describing the various shrinkage estimators in terms of the factorial experiment, examples of the applications of each are given. In order to evaluate the relative quality of the various estimators, Monte Carlo comparisons of the positive part versions of the shrinkage estimators are reported.

## 4.1 Applications of the Estimators

Recall from 1.1.3, that the cell means model for the two-way fixed effects factorial experiment, written in matrix notation is

$$\underline{y} = X\underline{\mu} + \underline{\varepsilon}\ ,\ \underline{\varepsilon} \sim N(\underline{0}, I\sigma^2) \qquad (4.1.1)$$

The OLS estimate for $\underline{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{ab})$ is

$$\hat{\underline{\mu}} = (\bar{y}_{1..}, \bar{y}_{2..}, \ldots, \bar{y}_{ab.})' \qquad (4.1.3)$$

where $\bar{y}_{ij.} = (1/n) \sum_{k=1}^{n} y_{ijk}$.

Often it is suspected that there will be no interaction between A and B. In terms of the factor effects model (1.1.1), this means it is suspected that $\gamma_{ij} = 0$ for all (i,j). Equivalently stated in terms of the cell means model, it is suspected that

$$R\,\underline{\mu} = \underline{0}\ , \text{ where } R = (\underline{-1}\ I_{(a-1)}) \otimes (\underline{-1}\ 1_{(b-1)}) \qquad (4.1.4)$$

This restriction can be restated as $R^*\underline{\mu} = \underline{0}$, where $R^*$ is an orthonormal version of R. The matrix $R^*$ will be of use in applying the empirical Bayes estimator later in this section. The restricted least squares estimate of $\underline{\mu}$ such that $R\underline{\mu} = 0$ is

$$\tilde{\underline{\mu}} = \hat{\underline{\mu}} - S^{-1} R' ( RS^{-1}R' )^{-1} ( R\,\hat{\underline{\mu}} ) \qquad (4.1.4)$$

To test $H_0$: $R\underline{\mu} = \underline{0}$ vs $H_1$: $R\underline{\mu} \neq \underline{0}$, an F test is used. The test statistic is

$$F = \frac{(R\hat{\underline{\mu}})'(RS^{-1}R')^{-1}(R\hat{\underline{\mu}})\ (N-ab)}{(\underline{y}-X\hat{\underline{\mu}})'(\underline{y}-X\hat{\underline{\mu}})\ (a-1)(b-1)} \qquad (4.1.5)$$

This F statistic will hence forth be written in a form better recognized by applied researchers, that is

$$F = MS_{AB}/MS_E,$$

where $MS_{AB}$ is the mean square attributed to the AB interaction in the ANOVA table, and $MS_E$ is the mean square due to experimental error. In general, F has the noncentral F distribution with $(a-1)(b-1)$ numerator degrees of freedom, $(N-ab)$ denominator degrees of freedom, and noncentrality parameter

$$\lambda = (R\underline{\mu})'(RS^{-1}R')^{-1}(R\underline{\mu})/(2\sigma^2) \qquad (4.1.6)$$

This is denoted

$$F \sim F(\lambda,(a-1)(b-1),(N-ab)).$$

The standard decision rule is reject $H_0$ if $F \geq c$, and accept $H_0$ if $F < c$, where $c=c(\alpha)$ is a constant chosen so that $Pr(F \geq c(\alpha) \mid H_0$ is true$)=\alpha$ (type 1 error rate).

4.1.1 Conventional Strategy Estimator

Being unsure of whether $H_0$ is true or false, the conventional strategy for estimating $\underline{\mu}$ may be stated as follows:

1) Test $H_0$: $R\underline{\mu} = \underline{0}$ using the F test.
2) If $H_0$ is rejected, then use $\hat{\underline{\mu}}$ as the estimator of $\underline{\mu}$
3) If $H_0$ is not rejected, then use $\tilde{\underline{\mu}}$ as the estimator of $\underline{\mu}$

The conventional strategy is itself an estimator, and will be called $\hat{\hat{\underline{\mu}}}$.

$$\hat{\underline{\mu}} = \begin{cases} \hat{\underline{\mu}} & \text{[illegible]} \\ \hat{\underline{\mu}} & \text{if the F-test accepts } H_0. \end{cases}$$

This can be rewritten as

$$\hat{\underline{\mu}} = \hat{\underline{\mu}} - I_{(0,c)}(F)\, S^{-1}R'(RS^{-1}R')^{-1}R\,\hat{\underline{\mu}}, \qquad (4.1.7)$$

where $I_{(0,c)}(F) = 1$ if $F < c$, $I_{(0,c)}(F) = 0$ if $F \geq c$, and $c = c(\alpha)$ is the $\alpha$ level critical value for $F$. In (4.1.7), $\hat{\underline{\mu}}$ is in the form of a preliminary test estimator as defined by Bock et al. (1973).

4.1.2 James Stein Estimators

Stein (1962) suggested that a shrinkage estimator might be used to shrink the estimates of the interaction parameters towards zero in a two-way factorial situation. Such an estimator is a special case of the estimators dealt with by Lee (1983). Both Stein and Lee deal with this problem assuming all $n_{ij}$'s are equal. The James-Stein estimator is now applied to the two-way fixed effects factorial experiment when not all the $n_{ij}$'s are equal.

Assume that the factorial experiment is such that $(a-1)(b-1) \geq 3$. Applying the James-Stein estimator stated in (3.4.12) to the cell means model, shrinking toward no interaction, the James-Stein estimator of $\underline{\mu}$ is

$$\tilde{\underline{\mu}}^S = (1 - c/F)\,\bar{\underline{\mu}} + (c/F)\,\hat{\underline{\mu}}, \qquad (4.1.8)$$

where $c = [(a-1)(b-1) - 2][N-ab]/[(a-1)(b-1)(N-ab+2)]$,

The James-Stein positive part estimator that compromises between $\bar{\underline{y}}$ and $\underline{\mu}$ as in (3.4.16) is

$$\hat{\underline{\mu}}^{S+} = \min\{1, c_1/F\}\,\tilde{\underline{\mu}} + (1-\min[1, c_1/F])\,\bar{\underline{y}} \quad , \qquad (4.1.10)$$

where $c_1$ is the same constant described in (3.4.16), and F is the F statistic stated in (4.1.4)

4.1.3 Empirical Bayes

The shrinkage estimator suggested by Efron and Morris (1973), was used as motivation for an empirical Bayes estimator in section 3.4. A specific application of the empirical Bayes estimator would be shrinking the estimates of the cell means towards no interaction in a two-way factorial experiment, with $(a-1)(b-1) \geq 3$

An empirical Bayes estimator for the cell means in a factorial experiment, shrinking towards no interaction (assuming $(a-1)(b-1) \geq 3$) is

$$\hat{\underline{\mu}}^{E} = (I - gM(\hat{v}(R^{*}S^{-1}R^{*\prime})^{-1} + 1)^{-1}R^{*})\,\hat{\underline{\mu}} \qquad (4.1.13)$$

where $\hat{v} = [(N-ab-2)\,F/\,(N-ab)-1]\,(a-1)(b-1)/tr((R^{*}S^{-1}R^{*\prime})^{-1})$ ,

$g=(N-ab-2)((a-1)(b-1)-2)/(N-ab+2)(a-1)(b-1)$ , F is the F test of $H_0$ $R^{*}\underline{\mu}=\underline{0}$ , and M and $S^{-1}$ are as in (3.4.46).

The matrix $R^{*}$ is the orthonormal version of R as in (3.4.46); $R^{*}$ has the property that $R^{*}R^{*\prime} = I$

The positive part empirical Bayes estimator of the cell means in the factorial experiment, shrinking towards no intersection is

$$\hat{\mu}_i = \begin{cases} [illegible] \\ \underline{\mu}_i & \text{if } \hat{\mu}_i^E \text{ violates } [\,\underline{\mu}_i\,,\,\bar{\mu}_i\,] \text{ toward } \bar{\mu}_i \\ \bar{\mu}_i & \text{if } \hat{\mu}_i^E \text{ violates } [\,\underline{\mu}_i\,,\,\bar{\mu}_i\,] \text{ toward } \bar{\mu}_i \end{cases} \qquad [illegible]$$

where $g = (N\text{-}eb\text{-}2)c_1/(N\text{-}eb)$. Recall that c_1 is the shrinkage constant used in (4.1.10) and (3.4.16), and M, $\bar{v} = R' S^{-1}$, and $\underline{u}$ are as in (4.1.13).

4.2 Examples

In this section the estimators described in section 4.1 will be applied in two examples. In Appendix B, computer programs written for SAS, PROC MATRIX to compute the restricted estimator,the James-Stein positive part estimator and empirical Bayes positive part estimator are given. Also a program is given which will help the user choose a shrinkage constant for the James-Stein estimator so that the estimator is minimax.

The checklist in Table 4.1 indicates the summary information needed to evaluate the estimators applied in the examples, using the computer programs.

These examples are based on this author's personal experiences involving agricultural experimentation while working for the agricultural consulting group of Institute of Food and Agricultural Sciences, University of Florida. The data given and the circumstances described, are contrived to illustrate computational techniques.

Table 4.1 Summary Information Required to Compute Estimators.

| | $\hat{\mu}$ | $n_{ij}$'s | a & b | $F_{AB}$ |
|---|---|---|---|---|
| Restricted | ✓ | ✓ | ✓ | |
| James-Stein + | ✓ | ✓ | ✓ | ✓ |
| Empirical Bayes + | ✓ | ✓ | ✓ | ✓ |

4.2.1 Example 1: Nematode Study

A 3×3 factorial experiment was conducted to determine the effect of factors A (nematocide) and B (fertilizer) on the number of nematodes found in the soil around citrus trees. Three different nematocides were compared. Also three fertilizers were being evaluated. The experimental units were potted orange trees in a large citrus nursery. The application of the nine factorial combinations of A and B was applied randomly, so that each factorial combination was applied to twelve trees.

Three months after the treatments were applied, a soil sample was taken from each potted tree. A count was taken of the nematodes found in each soil sample. The response variable, y, is the square root of the number of nematodes in a soil sample.

Unfortunately, several trees were destroyed following the application of the treatments, and before soil samples were obtained, due to a citrus canker eradication program. In Table 4.2, the observed responses are listed for each tree. Note there is not equal replication for each factorial combination. Using SAS, an analysis of variance was performed. The ANOV results are shown in Table 4.3.

[illegible] Number of Observed Nematodes for Each [illegible] [illegible]lizer Combination

| | | B | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | | 7.8740 | 5.5678 | 6.7823 |
| | | 4.6590 | 3.0000 | 7.7460 |
| | | 8.6603 | 5.0000 | 3.7417 |
| | | 7.5498 | 3.8730 | 4.3589 |
| | | 9.7580 | 5.3852 | 8.1240 |
| | | 4.5826 | 5.7446 | |
| | | 9.3274 | 7.8811 | |
| | | 7.4162 | | |
| | | 3.8730 | | |
| | | 9.0000 | | |
| A | | 10.0995 | 8.6603 | 3.0000 |
| | | 8.8318 | 7.7460 | 9.4340 |
| | | 9.7980 | 7.8811 | 2.0000 |
| | | 4.8590 | 4.8590 | 9.3274 |
| | | 9.3808 | 8.7750 | 4.5826 |
| | | 7.4162 | 5.0000 | 6.0000 |
| | | 9.4340 | 9.2195 | 5.2917 |
| | | 9.0554 | 5.6569 | |
| | | 4.0000 | 7.1414 | |
| | | | 4.6904 | |
| | | | 8.4853 | |
| | | 10.6301 | 6.5574 | 10.4881 |
| | | 10.2470 | 3.4641 | 9.7468 |
| | | 8.6023 | 6.6332 | 5.1962 |
| | | 10.8187 | 8.2462 | 9.2736 |
| | | 11.7047 | 9.5917 | 8.9443 |
| | | 6.4807 | 9.5917 | 10.0000 |
| | | 11.0905 | 10.6771 | 8.8318 |
| | | 10.8167 | 8.4853 | 10.1489 |
| | | 9.6954 | 10.0000 | 12.0416 |
| | | 5.4772 | 8.1240 | 9.1652 |
| | | | | 7.7460 |
| | | | | 6.8557 |

[illegible] Experiment

| [illegible] | [illegible] | [illegible] | [illegible] |
|---|---|---|---|
| A | [illegible] | 52.335 | |
| B | [illegible] | 19.180 | |
| AB | [illegible] | 4.434 | [illegible].05 (pval = .3853) |
| error | [illegible] | 4.204 | |

The cell averages are listed in Table 4.4 as well as the restricted least squares estimate of cell means, assuming no interaction. A preliminary test strategy using $\alpha = .25$ would choose the restricted estimator in this example. That is reflected in Table 4.4. Also, the positive part versions of the compromise estimators being studied are listed in Table 4.4. These estimators were computed using the computer programs listed in Appendix B.

For the James-Stein positive part estimator, the choice of shrinkage constant was made using the computer program 2, listed in Appendix B. The choice of $c_1 = .474837$ insures that the James-Stein positive part estimator will be minimax for estimating cell means. Also, Table 4.4 lists the various estimates of marginal means.

In Figures 4.1 and 4.2, geometric interpretations of the compromise made using the shrinkage estimators are displayed. Note that every element of the James-Stein positive part estimator is shrunk by the same amount, from the OLS estimator toward the restricted estimator. The elements of the empirical Bayes positive part estimator are all adjusted by different amounts. The idea of graphing the estimators in this way comes from Casella (1985).

[illegible] and Marginal Means Using Various [illegible]

| [illegible] | [illegible] | Restricted | PTE $\alpha = 25$ | J-S+ $\alpha_1 = 4748$ | ZB+ |
|---|---|---|---|---|---|
| 1,1 | 7.289 | 7.109 | 7.109 | 7.206 | 7.206 |
| 1,2 | 5.179 | 5.666 | 5.666 | 5.399 | 5.405 |
| 1,3 | 6.151 | 5.828 | 5.828 | 6.005 | 5.999 |
| 2,1 | 8.102 | 7.976 | 7.976 | 8.045 | 8.043 |
| 2,2 | 7.087 | 6.533 | 6.533 | 6.836 | 6.830 |
| 2,3 | 5.662 | 6.695 | 6.695 | 6.129 | 6.141 |
| 3,1 | 9.556 | 9.849 | 9.849 | 9.689 | 9.692 |
| 3,2 | 8.137 | 8.406 | 8.406 | 8.259 | 8.261 |
| 3,3 | 9.036 | 8.568 | 8.568 | 8.824 | 8.820 |
| | | | Marginal Means Estimates | | |
| A | | | | | |
| 1 | 6.209 | 6.201 | 6.201 | 6.204 | 6.202 |
| 2 | 6.950 | 7.068 | 7.068 | 7.003 | 7.005 |
| 3 | 8.910 | 8.941 | 8.941 | 8.924 | 8.924 |
| B | | | | | |
| 1 | 8.319 | 8.314 | 8.314 | 8.314 | 8.314 |
| 2 | 6.801 | 6.868 | 6.868 | 6.831 | 6.832 |
| 3 | 6.950 | 7.031 | 7.031 | 6.986 | 6.987 |

In Figure 4.3, the impact that compromise estimators can have on the interpretation of results is illustrated. Note that the OLS estimator indicates nematocide 2 may be the best choice for minimizing nematodes, in a grove where fertilizer 3 is used. The shrinkage estimators indicate that nematocide 1 may be the best choice (as does the restricted estimator). On the other hand, the restricted estimator indicates that fertilizer 2 is the best choice for minimizing nematodes.

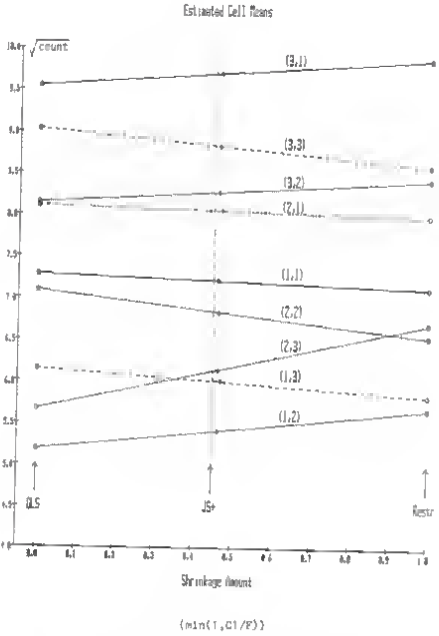



Figure 4.1 Geometric Interpretation of James-Stein Estimate of Cell Means Using Nematode Example Diamond is Empirical Bayes estimator

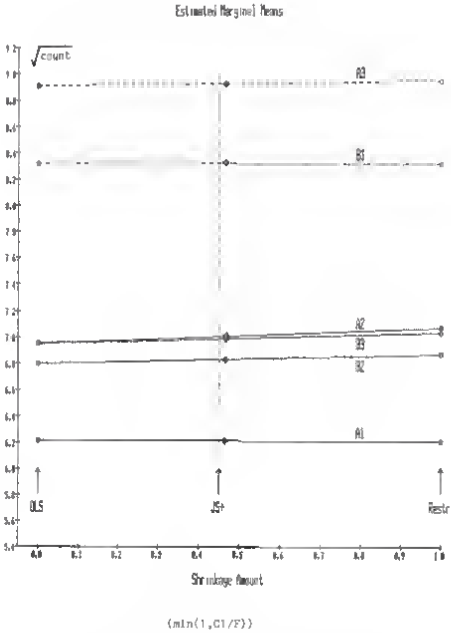
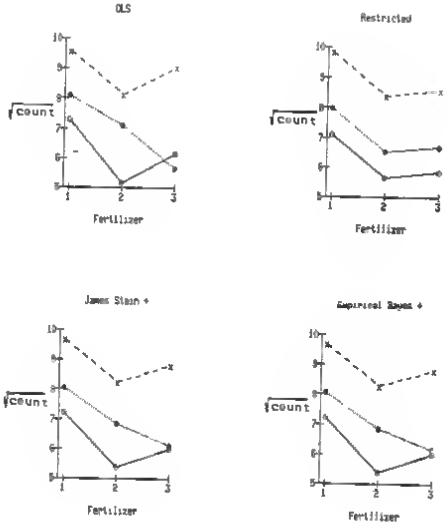



Figure 4.2 Geometric Interpretation of James-Stein Estimate of Marginal Means Using Nematode Study. Diamond is Empirical Bayes estimator.




Legend: ⊗ = Nematocide 1, ● = Nematocide 2, ▾ = Nematocide 3

Figure # 3 Impact on Compromise Estimators on Interpretation of Results in Nematode Example

in a grove where nematocide 1 is being used. The compromise estimators indicate that fertilizer 3 may be best, in groves using nematocide 2 (as does the OLS estimator).

4.2.2 Example 2

The data from a 4x2 factorial experiment is summarized in Table 4.5. The summary information will be used to estimate cell means and marginal means, using the various estimators, for this experiment. Additionally, MSE = 1 and the interaction F statistic if computed will be F= 77 (pval=.53).

Table 4.5 Summary Information for Example 2

Number of Observations Per Factor Level Combination

| | | B: 1 | B: 2 |
|---|---|---|---|
| A | 1 | 3 | 1 |
| | 2 | 3 | 1 |
| | 3 | 1 | 10 |
| | 4 | 3 | 1 |

Cell Averages

| | | B: 1 | B: 2 |
|---|---|---|---|
| A | 1 | 9 | 9 |
| | 2 | 10 | 9 |
| | 3 | 10 | 11 |
| | 4 | 11 | 10 |

The OLS estimator of the cell means is simply the vector of cell averages. Table 4.6 gives estimates of the cell means using the various estimation strategies described in this chapter. Notice that using an $\alpha$=.25 error rate in the preliminary test of interaction will cause the preliminary test estimator to take the values of the restricted estimator in this example. Also, the shrinkage constant for the

Table 4.3 Predicted Cell Means and Marginal Means Using Various Estimators:

| Cell | OLS | Restricted | PTE $\theta = .25$ | J-S+ $c_1 = .2941$ | BR |
|---|---|---|---|---|---|
| | | Cell Means Estimators | | | |
| 1,1 | 9 | 9.047 | 9.047 | 9.018 | 9.014 |
| 1,2 | 5 | 8.860 | 8.860 | 8.546 | 8.958 |
| 1,3 | 10 | 9.797 | 9.797 | 9.922 | 9.921 |
| 2,1 | 5 | 9.610 | 9.610 | 9.233 | 9.238 |
| 2,2 | 10 | 11.079 | 11.079 | 10.412 | 10.433 |
| 2,3 | 11 | 10.892 | 10.892 | 10.959 | 10.957 |
| 3,1 | 11 | 10.797 | 10.797 | 10.922 | 10.921 |
| 3,2 | 10 | 10.610 | 10.610 | 10.233 | 10.238 |
| | | Marginal Means Estimates | | | |
| A | | | | | |
| 1 | 9.0 | 8.953 | 8.953 | 8.982 | 8.986 |
| 2 | 9.5 | 9.703 | 9.703 | 9.578 | 9.579 |
| 3 | 10.5 | 10.986 | 10.986 | 10.686 | 10.695 |
| 4 | 10.5 | 10.703 | 10.703 | 10.578 | 10.579 |
| B | | | | | |
| 1 | 10 | 10.180 | 10.180 | 10.069 | 10.072 |
| 2 | 9.75 | 9.893 | 9.893 | 9.843 | 9.847 |

James-Stein positive part estimator is chosen to be $c_1 = .2941$, based on the results of applying program 2 in Appendix E

As in example 1, a geometric interpretation of the compromise estimators is shown in Figure 4.4. Also, in Figure 4.5, the graph of the predicted response versus factor A and factor B is displayed

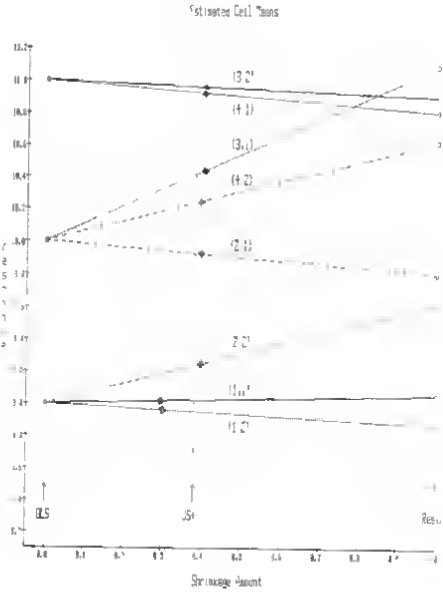



Figure 4.4 Geometric Interpretation of James-Stein Estimate of Cell Means Using Example 2 Data. Diamond is Empirical Bayes estimator.

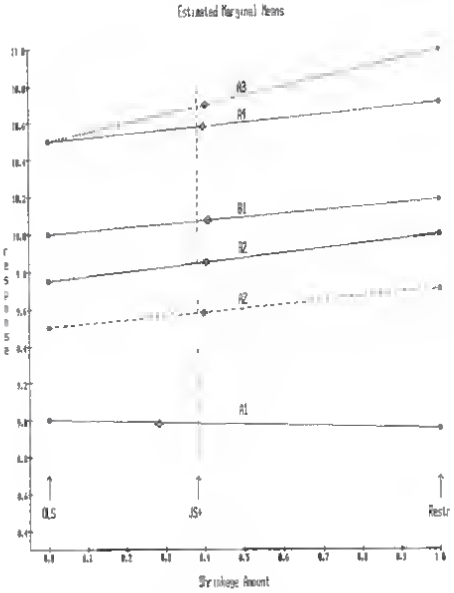
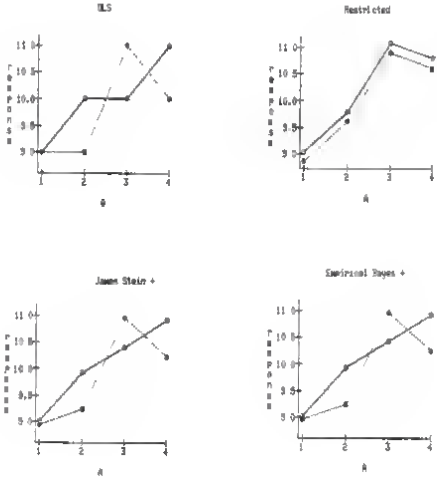



Figure 4 5 Geometric Interpretation of James-Stein Estimate of Marginal Means Using Example 2. Diamond is Empirical Bayes estimator.




Legend: o B=1, • B=2

Figure 4.6 Impact of Compromise Estimators on Interpretation of Results in Example 2.

## 4.3 Monte Carlo Comparisons

In section 3.2 risk functions $R_1(\underline{e},\underline{\mu})$, and $R_2(\underline{e},\underline{\mu})$ were [illegible] where $\underline{e}$ is some estimator of $\underline{\mu}$. Recall that $R_1(\underline{e},\underline{\mu})$ is the expected squared distance of $\underline{e}$ from the true cell means. Also $R_2(\underline{e},\underline{\mu})$ is the expected squared error of the estimate of marginal means that results from using the estimator $\underline{e}$ as an estimator of the cell means. In this section simulation techniques are described that were used to estimate the difference in risk incurred by using one of the biased estimators of the cell means, rather than the unbiased OLS estimator.

### 4.3.1 Simulation Description

Using Monte Carlo techniques, two quantities were estimated

$$D_1(\underline{e}) = R_1(\underline{e},\underline{\mu}) - R_1(\hat{\underline{\mu}},\underline{\mu}) \qquad (4.3.1)$$

and

$$D_2(\underline{e}) = R_2(\underline{e},\underline{\mu}) - R_2(\hat{\underline{\mu}},\underline{\mu}),$$

where $\hat{\underline{\mu}}$ is the OLS estimator.

The estimators under consideration were the restricted least squares estimator ($\tilde{\underline{\mu}}$), a preliminary test estimator ($\hat{\underline{\mu}}$), a James-Stein positive part estimator ($\hat{\underline{\mu}}^{S+}$), and an Empirical Bayes positive part estimator ($\hat{\underline{\mu}}^{E+}$). The preliminary test estimator considered was one that tests for interaction at the $\alpha = .25$ level. The James-Stein positive part estimator considered uses the shrinkage constant

$$c_1 = [(a-1)(b-1)-2][N-ab]/[(a-1)(b-1)(N-ab+2)] \qquad (4.3.2)$$

The Empirical Bayes positive part estimator under consideration uses

$$g = [(a-1)(b-1)-2][N-ab-2]/[(a-1)(b-1)(N-ab+2)], \qquad (4.3.5)$$

as the shrinkage constant. Recall from previous discussion that the choice of g given in (4.3.3) will cause the Empirical Bayes positive part estimator to emulate the James-Stein positive part estimator with $c_1$ as in (4.3.2) in that special case that the cell sizes of the factorial experiment are all equal.

Two situations were considered: an unbalanced 3×3 factorial experiments, and a balanced 3×3 factorial experiment. Both of these situations were constrained to have N=27 total observations in the factorial experiment. In Table 4.7 the number of observations per cell for each situation is given.

Table 4.7 Cell Sizes Considered in the Simulations.

Situation 1

| | | B | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| A | 1 | 10 | 5 | 1 |
| | 2 | 2 | 1 | 1 |
| | 3 | 1 | 5 | 1 |

Situation 2

| | | B | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| A | 1 | 3 | 3 | 3 |
| | 2 | 3 | 3 | 3 |
| | 3 | 3 | 3 | 3 |

In situation 1, the pattern of interaction, [illegible] as the amount of interaction that exists in $\underline{y}$ has an influence on the risk incurred by using the biased estimators being studied. Therefore, eight different patterns of interaction were considered for situation 1. These patterns of interaction are given in Table 4.8. In Appendix A, details about these choices of interaction pattern are described. At this point it suffices to say that these patterns of interaction are representative of the population of all possible interaction patterns.

For each pattern of interaction, amounts of interaction corresponding to

$$\lambda = 0, .5, 1, 1.5, 2, 3, 12.5, 5, 12, \text{ and } 8$$

were considered. Recall that $\lambda$ is the interaction F statistic's noncentrality parameter as in (4.1.5).

In situation 2, the pattern of interaction has no effect on risk, so only one pattern of interaction was considered. That pattern of interaction is given in Table 4.8. The same amounts of interaction were considered in situation 2 as were in situation 1.

The number of simulations used for each situation was 200; in each simulation, data were generated to correspond to the cell sizes given in Table 4.7. The data were generated using the N(0,1) random number generator called RANNOR, in SAS. Onto this random error, interaction was added corresponding to each pattern and amount of interaction being considered.

Table 4.8 Patterns of Interaction Used in Simulation.

Situation 1: Unbalanced Case

| Cell (i,j) | Interaction Pattern 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1.1 | -0.037685 | -0.092701 | -0.099569 | -0.011815 | -0.090312 | 0.011611 | 0.084163 | 0.090869 |
| 1.2 | 0.003710 | 0.169542 | 0.208525 | 0.030233 | 0.169045 | -0.015585 | -0.050471 | 0.043520 |
| 1.3 | 0.358297 | 0.079295 | -0.048438 | -0.033019 | 0.057896 | -0.036182 | -0.589270 | -0.226191 |
| 2.1 | -0.067178 | 0.307044 | 0.285100 | -0.054490 | 0.083772 | -0.263267 | -0.194647 | -0.096247 |
| 2.2 | -0.179209 | -0.570583 | -0.354904 | -0.488042 | 0.024073 | -0.073861 | -0.104042 | -0.477392 |
| 2.3 | 0.313565 | 0.054495 | -0.215297 | 0.597023 | -0.201616 | 0.600395 | 0.493335 | 0.669886 |
| 3.1 | 0.511203 | 0.312917 | 0.425488 | 0.227125 | 0.735576 | 0.410427 | -0.452332 | 0.153950 |
| 3.2 | 0.032132 | -0.035425 | -0.137844 | 0.067376 | -0.175859 | 0.030757 | 0.071280 | 0.051959 |
| 3.3 | -0.671861 | -0.135791 | 0.263735 | -0.564004 | 0.143720 | -0.564213 | 0.094934 | -0.443695 |

Situation 2: Balanced Case

| Cell (i,j) | |
|---|---|
| 1.1 | 0.226962 |
| 1.2 | -0.034227 |
| 1.3 | -0.192755 |
| 2.1 | -0.022094 |
| 2.2 | 0.179683 |
| 2.3 | -0.15759 |
| 3.1 | -0.204888 |
| 3.2 | -0.145456 |
| 3.3 | 0.350344 |

Based on Lemma 3.4.1 it was unnecessary to consider different amounts of main effects in this simulation. From Lemma 3.4.1, estimators of the form under consideration in this chapter differ in risk only according to the estimate of the interaction that they employ. Therefore without loss of generality, the main effects were allowed to be zero. Also, $\sigma^2$ was taken to be one without loss of generality.

For each simulation, and for each pattern and amount of interaction, the various estimators were applied, using SAS / PROC MATRIX. For each simulation×pattern×amount×estimator combination, the following two quantities were computed:

$$L_1(\underline{e}) = (\underline{e} - \underline{\mu})'(\underline{e} - \underline{\mu})$$

and (4.3.4)

$$L_2(\underline{e}) = (\underline{e} - \underline{\mu})'D(\underline{e} - \underline{\mu})$$

Next the differences

$$d_1(\underline{e}) = L_1(\underline{e}) - L_1(\hat{\underline{\mu}})$$

and (4.3.5)

$$d_2(\underline{e}) = L_2(\underline{e}) - L_2(\hat{\underline{\mu}})$$

were computed for each simulation×pattern×amount×estimator.

Estimates of $D_1(\underline{e})$ and $D_2(\underline{e})$ were obtained for each pattern×amount×estimator by averaging the $d_1(\underline{e})$'s and $d_2(\underline{e})$'s respectively, over all 200 simulations. Also the standard errors of the

estimators were estimated. In Tables [illegible] through 4.16 the results for situation 1 are tabulated. Results for situation 2 are reported in Table 4.17.

4.3.2 Conclusions

Using RSI, $D_1(\underline{e})$ and $D_2(\underline{e})$ were charted against the noncentrality parameter, $\lambda$. In Figure 4.7 the charts for $D_1(\underline{e})$ are displayed for each interaction pattern in the unbalanced case. Similarly, Figure 4.8 displays the charts for $D_2(\underline{e})$ for each interaction pattern in the unbalanced case.

Only the chart of $D_1(\underline{e})$ versus lambda (Figure 4.9) is required for the balanced data situation. This is because $D_2(\underline{e})$ is zero for all estimators, when the numbers of observations per cell are equal. Also it can be noted in Figure 4.9 or Table 4.11 that the Empirical Bayes positive part estimator has exactly the same risk as the James-Stein positive part estimator when cell sizes are equal.

Based on the two situations considered, the two shrinkage estimators tended to have less risk in estimating cell or marginal means as compared to the OLS and preliminary test estimators. Admittedly the gains of the shrinkage estimators were in some cases modest. Even so the practicing data analyst generally does not know the pattern of interaction a priori. At any time, the pattern of interaction and the pattern of unbalance in an experiment may be such that the shrinkage estimators provided substantial improvement over OLS and preliminary test estimators.

The fact that shrinkage estimators have smaller risks is of little value to the data analyst, unless this property can be capitalized on by

Table 4-9 Risk Increase ($D_1$) of Restricted Estimator over OLS Estimator, Situation 1, Table 4-7. Risk increase in units.

| Interaction Pattern | | Interaction Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 0 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -3.34 | -2.37 | -1.40 | -0.42 | 0.55 | 2.75 | 6.65 | 12.28 |
| | std* | .174 | .177 | .179 | .181 | .183 | .186 | .191 | .198 |
| 2 | mean | -3.34 | -2.65 | -1.93 | -1.25 | -0.35 | 1.04 | 3.85 | 7.92 |
| | std | .174 | .181 | .186 | .191 | .197 | .207 | .225 | .249 |
| 3 | mean | -3.34 | -2.77 | -2.20 | -1.62 | -1.04 | .26 | 2.36 | 5.89 |
| | std | .174 | .180 | .187 | .194 | .200 | .215 | .238 | .269 |
| 4 | mean | -3.34 | -2.38 | -1.42 | 0.45 | 0.52 | 2.71 | 6.59 | 12.19 |
| | std | .174 | .178 | .181 | .183 | .185 | .189 | .195 | .204 |
| 5 | mean | -3.34 | -2.66 | -1.98 | -1.29 | -0.61 | 0.92 | 3.65 | 7.57 |
| | std | .174 | .179 | .185 | .191 | .197 | .210 | .232 | .261 |
| 6 | mean | -3.35 | -2.42 | -1.50 | -0.57 | 0.35 | 2.43 | 6.12 | 11.44 |
| | std | .174 | .177 | .178 | .180 | .182 | .185 | .191 | .198 |
| 7 | mean | -3.34 | -2.48 | -1.61 | -0.74 | 0.13 | 2.08 | 5.55 | 10.55 |
| | std | .174 | .177 | .180 | .182 | .185 | .191 | .201 | .214 |
| 8 | mean | -3.34 | -2.38 | -1.42 | -0.45 | 0.52 | 2.71 | 6.58 | 12.17 |
| | std | .174 | .177 | .179 | .181 | .183 | .186 | .192 | .200 |

* Note that std is the standard error of the mean

Table 4.7. Risk Increase (D) [illegible] over OLS Estimator, Situation 1: Risk increase in $\sigma^2$ units.

| Iteration Pattern | | Interaction Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -2.07 | -0.91 | -0.22 | 0.34 | 0.66 | 0.99 | 0.78 | 0.27 |
| | std* | .136 | 109 | .095 | .095 | .102 | .136 | 163 | 121 |
| 2 | mean | -2.07 | -1.08 | -0.51 | -0.05 | 0.22 | 0.37 | 0.46 | 0.20 |
| | std | .136 | 118 | .105 | .096 | .079 | .083 | .103 | 085 |
| 3 | mean | -2.07 | -1.18 | -0.63 | -0.21 | -0.04 | 0.17 | 0.31 | 0.16 |
| | std | .136 | 123 | 113 | .091 | .078 | .070 | .081 | .069 |
| 4 | mean | -2.07 | -1.07 | -0.30 | 0.28 | 0.62 | 0.96 | 0.65 | 0.13 |
| | std | .136 | .121 | 105 | 093 | 100 | .136 | .145 | .078 |
| 5 | mean | -2.07 | -1.12 | -0.44 | -0.04 | 0.20 | 0.40 | 0.45 | 0.25 |
| | std | .136 | .127 | .106 | .091 | .083 | .079 | .103 | .096 |
| 6 | mean | -2.07 | -0.97 | -0.27 | 0.22 | 0.60 | 1.00 | 0.65 | 0.14 |
| | std | 136 | .111 | 093 | .091 | .097 | .126 | .143 | 084 |
| 7 | mean | -2.07 | -1.09 | -0.38 | 0.01 | 0.29 | 0.71 | 0.49 | 0.23 |
| | std | 136 | .123 | .104 | .099 | .099 | .123 | .120 | .104 |
| 8 | mean | -2.07 | -1.00 | -0.25 | 0.28 | 0.61 | 0.89 | 0.73 | 0.19 |
| | std | .136 | 117 | 101 | 094 | 099 | 131 | 151 | .094 |

* Note that std is the standard error of the mean.

[illegible] Increase (D [illegible]) of James-Stein [illegible] over OLS Estimator, Situation [illegible], Table 4.[illegible] [illegible] increase in $\sigma^2$ units.

| Interaction Pattern | | Interaction Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -1.94 | -1.45 | -1.08 | -0.81 | -0.60 | -0.35 | -0.18 | -0.14 |
| | std* | .068 | .076 | .080 | .083 | .087 | .086 | .076 | .065 |
| 2 | mean | -1.94 | -1.55 | -1.28 | -1.06 | -0.89 | -0.63 | -0.41 | -0.28 |
| | std | .068 | .075 | .073 | .071 | .070 | .066 | .056 | .048 |
| 3 | mean | -1.94 | -1.59 | -1.34 | -1.14 | -0.98 | -0.72 | -0.48 | -0.34 |
| | std | .068 | .072 | .070 | .068 | .067 | .062 | .055 | .044 |
| 4 | mean | -1.94 | -1.46 | -1.11 | -0.86 | -0.66 | -0.41 | -0.22 | -0.14 |
| | std | .068 | .077 | .080 | .079 | .080 | .079 | .069 | .057 |
| 5 | mean | -1.94 | -1.55 | -1.27 | -1.05 | -0.86 | -0.61 | -0.39 | -0.26 |
| | std | .068 | .072 | .072 | .072 | .073 | .070 | .062 | .052 |
| 6 | mean | -1.94 | -1.48 | -1.11 | -0.84 | -0.66 | -0.40 | -0.20 | -0.10 |
| | std | .068 | .075 | .080 | .082 | .080 | .075 | .073 | .062 |
| 7 | mean | -1.94 | -1.54 | -1.21 | -0.96 | -0.78 | -0.51 | -0.25 | -0.13 |
| | std | .068 | .075 | .079 | .078 | .075 | .071 | .075 | .066 |
| 8 | mean | -1.94 | -1.47 | -1.12 | -0.87 | -0.68 | -0.41 | -0.20 | -0.12 |
| | std | .068 | .077 | .080 | .080 | .080 | .079 | .073 | .060 |

* Note that std is the standard error of the mean.

Table [illegible] Increase ($D_T$) of Empirical Bayes + Estimator over OLS Estimator, Situation 1, Table 4.7. Risk increase in $\sigma^2$ units.

| Intercept Pattern | | Interaction Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -1.87 | -1.43 | -1.16 | -0.94 | -0.78 | -0.42 | -0.17 | -0.02 |
| | std* | .077 | .080 | .077 | .074 | .079 | .082 | .081 | .078 |
| 2 | mean | -1.87 | -1.50 | -1.22 | -1.10 | -0.99 | -0.67 | -0.49 | -0.36 |
| | std | .077 | .084 | .081 | .076 | .071 | .075 | .060 | .050 |
| 3 | mean | -1.87 | -1.49 | -1.27 | -1.15 | -1.05 | -0.79 | -0.56 | -0.45 |
| | std | .077 | .080 | .078 | .073 | .064 | .072 | .059 | .041 |
| 4 | mean | -1.87 | -1.44 | -1.18 | -0.95 | -0.77 | -0.47 | -0.21 | -0.05 |
| | std | .077 | .079 | .078 | .075 | .078 | .075 | .073 | .073 |
| 5 | mean | -1.87 | -1.46 | -1.27 | -1.10 | -0.98 | -0.64 | -0.46 | -0.38 |
| | std | .077 | .080 | .078 | .076 | .073 | .083 | .078 | .054 |
| 6 | mean | -1.87 | -1.48 | -1.18 | -0.96 | -0.78 | -0.46 | -0.20 | -0.01 |
| | std | .077 | .079 | .080 | .078 | .078 | .077 | .074 | .077 |
| 7 | mean | -1.87 | -1.54 | -1.28 | -1.04 | -0.85 | -0.60 | -0.33 | -0.14 |
| | std | .077 | .074 | .072 | .077 | .077 | .071 | .068 | .076 |
| 8 | mean | -1.87 | -1.47 | -1.21 | -0.96 | -0.74 | -0.46 | -0.20 | -0.03 |
| | std | .077 | .078 | .076 | .78 | .079 | .075 | .074 | .074 |

* Note that std is the standard error of the mean

[illegible] increase ($D_2$ [illegible] estimator over OLS estimator. Situation [illegible]. Risk increase in $\sigma^2$ units

| Intersection Pattern | | Intersection Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -.270 | -.235 | -.200 | -.165 | -.131 | -.053 | .086 | .286 |
| | std* | .0324 | .0341 | .0353 | .0364 | .0374 | .0395 | .0429 | .0472 |
| 2 | mean | -.270 | -.197 | -.121 | -.045 | .032 | .205 | .513 | .859 |
| | std | .0324 | .0375 | .0409 | .0439 | .0466 | .0521 | .0604 | .0704 |
| 3 | mean | -.270 | -.174 | -.080 | .014 | .108 | .320 | .694 | 1.235 |
| | std | .0324 | .0382 | .0426 | .0465 | .0500 | .0571 | .0678 | .0806 |
| 4 | mean | -.270 | -.248 | -.223 | -.197 | -.171 | -.111 | -.003 | .154 |
| | std | .0324 | .0344 | .0357 | .0369 | .0380 | .0403 | .0440 | .0486 |
| 5 | mean | -.270 | -.143 | -.022 | .098 | .217 | .485 | .859 | 1.64 |
| | std | .0324 | .0389 | .0444 | .0452 | .0536 | .0623 | .0753 | .0908 |
| 6 | mean | -.270 | -.256 | -.243 | -.231 | -.218 | -.189 | -.137 | -.064 |
| | std | .0324 | .0330 | .0338 | .0346 | .0354 | .0372 | .0401 | .0440 |
| 7 | mean | -.270 | -.193 | -.112 | -.031 | .051 | .236 | .564 | 1.040 |
| | std | .0324 | .0341 | .0362 | .0384 | .0404 | .0446 | .0517 | .0604 |
| 8 | mean | -.270 | -.250 | -.226 | -.202 | -.177 | -.121 | -.019 | .128 |
| | std | .0324 | .0340 | .0352 | .0363 | .0373 | .0395 | .0430 | .0474 |

* Note that std is the standard error of the mean.

[illegible] Degrees ($D_2$ [illegible] Situation 1, [illegible] 7. Risk [illegible]

| Interaction Pattern | | Interaction Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -.172 | -.102 | -.057 | -.015 | -.002 | .025 | .021 | .009 |
| | std* | .0263 | .0231 | .0214 | .0197 | .0177 | .0131 | .0093 | .0065 |
| 2 | mean | -.172 | -.069 | -.017 | .037 | .055 | .051 | .052 | .024 |
| | std | .0263 | .0224 | .0219 | .0223 | .0179 | .0179 | .0141 | .0100 |
| 3 | mean | -.172 | -.056 | .012 | .062 | .056 | .073 | .066 | .030 |
| | std | .0263 | .0234 | .0240 | .0211 | .0182 | .0192 | .0171 | .0127 |
| 4 | mean | -.172 | -.118 | -.082 | -.045 | -.081 | .010 | .000 | .000 |
| | std | .0263 | .0235 | .0208 | .0176 | .0149 | .0136 | .0085 | .0021 |
| 5 | mean | -.172 | -.051 | .057 | .103 | .118 | .120 | .097 | .043 |
| | std | .0263 | .0256 | .0238 | .0259 | .0250 | .0234 | .0238 | .0179 |
| 6 | mean | -.172 | -.107 | -.079 | -.065 | -.047 | -.022 | .000 | .004 |
| | std | .0263 | .0214 | .0203 | .0195 | .0173 | .0123 | .0077 | .0034 |
| 7 | mean | -.172 | -.089 | -.032 | -.004 | .026 | .082 | .047 | .028 |
| | std | .0263 | .0241 | .0218 | .0219 | .0205 | .0205 | .0150 | .0130 |
| 8 | mean | -.172 | -.114 | -.076 | -.057 | -.027 | .005 | -.006 | .000 |
| | std | .0263 | .0232 | .0197 | .0183 | .0146 | .0130 | .0082 | .0021 |

* Note that std is the standard error of the mean.

Table 4.15 Risk Increase ($D_2$) of James-Stein + Estimator over OLS Estimator, Situation 1, Table 4.7; Risk increase in $\sigma^2$ units.

| Intersection Pattern | | Intersection Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -.153 | -.126 | -.105 | -.089 | -.077 | -.059 | -.042 | -.030 |
| | std* | .0148 | .0141 | .0129 | .0119 | .0111 | .0094 | .0074 | .0057 |
| 2 | mean | -.153 | -.161 | -.094 | -.076 | -.062 | -.042 | -.024 | -.015 |
| | std | .0148 | .0153 | .0148 | .0140 | .0133 | .0116 | .0098 | .0084 |
| 3 | mean | -.153 | -.108 | -.081 | -.061 | -.044 | -.022 | -.006 | -.002 |
| | std | .0148 | .0154 | .0149 | .0145 | .0142 | .0132 | .0119 | .0099 |
| 4 | mean | -.153 | -.133 | -.118 | -.106 | -.095 | -.076 | -.055 | -.039 |
| | std | .0148 | .0138 | .0126 | .0116 | .0107 | .0090 | .0070 | .0054 |
| 5 | mean | -.153 | -.094 | -.060 | -.035 | -.018 | -.006 | -.015 | .015 |
| | std | .0148 | .0160 | .0160 | .0163 | .0168 | .0158 | .0139 | .0116 |
| 6 | mean | -.153 | -.134 | -.119 | -.107 | -.096 | -.078 | -.055 | -.039 |
| | std | .0148 | .0138 | .0126 | .0114 | .0104 | .0088 | .0070 | .0052 |
| 7 | mean | -.153 | -.123 | -.093 | -.072 | -.056 | -.032 | -.011 | -.004 |
| | std | .0148 | .0148 | .0149 | .0141 | .0133 | .0121 | .0119 | .0096 |
| 8 | mean | -.153 | -.134 | -.119 | -.107 | -.098 | -.079 | -.057 | -.040 |
| | std | .0148 | .0135 | .0125 | .0116 | .0107 | .0090 | .0068 | .0052 |

* Note that std is the standard error of the mean

Table 4.16 Risk Increase ($D_2$) of Empirical Bayes + Estimator over [illegible] Estimator, Situation 1, Table 4.7 Risk Increase in [illegible] units.

| Interaction Pattern | | Interaction Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | mean | -.145 | -.103 | -.092 | -.084 | -.068 | -.055 | -.038 | -.023 |
| | std* | .0155 | .0153 | .0133 | .0123 | .0128 | .0094 | .0078 | .0065 |
| 2 | mean | -.145 | -.098 | -.074 | -.074 | -.065 | -.027 | -.024 | -.020 |
| | std | .0155 | .0169 | .0166 | .0139 | .0130 | .0125 | .0092 | .0075 |
| 3 | mean | -.145 | -.083 | -.064 | -.056 | -.050 | -.019 | -.007 | -.010 |
| | std | .0155 | .0168 | .0154 | .0145 | .0125 | .0139 | .0104 | .0077 |
| 4 | mean | -.145 | -.105 | -.102 | -.089 | -.081 | -.072 | -.054 | -.036 |
| | std | .0155 | .0151 | .0137 | .0126 | .0112 | .0094 | .0079 | .0063 |
| 5 | mean | -.145 | -.066 | -.052 | -.033 | -.026 | .017 | .019 | .006 |
| | std | .0155 | .0175 | .0173 | .0167 | .0156 | .0177 | .0153 | .0101 |
| 6 | mean | -.145 | -.119 | -.098 | -.099 | -.101 | -.077 | -.054 | -.036 |
| | std | .0155 | .0150 | .0147 | .0128 | .0115 | .0092 | .0073 | .0061 |
| 7 | mean | -.145 | -.122 | -.095 | -.067 | -.060 | -.034 | -.014 | .002 |
| | std | .0155 | .0140 | .0129 | .0148 | .0144 | .0121 | .0103 | .0107 |
| 8 | mean | -.145 | -.119 | -.108 | -.095 | -.078 | -.073 | -.053 | -.036 |
| | std | .0155 | .0143 | .0141 | .0131 | .0123 | .0094 | .0074 | .0062 |

* Note that std is the standard error of the mean.

[illegible] Risk Increase ($D_1$) of the Biased Estimators over OLS Estimator, Situation 2, Table 4.7. Risk increase in $e^2$ units.

| Estimator | Intersection Noncentrality Parameter 0 | .5 | 1 | 1.5 | 2 | 3.125 | 5.12 | 8 |
|---|---|---|---|---|---|---|---|---|
| Restricted | | | | | | | | |
| mean | -1.34 | -1.01 | -0.67 | -0.34 | -0.01 | .74 | 2.07 | 3.99 |
| std* | .067 | .067 | .067 | .067 | .067 | .067 | .067 | .067 |
| FTE | | | | | | | | |
| mean | -0.70 | -0.36 | -0.10 | 0.06 | 0.13 | 0.20 | 0.20 | 0.08 |
| std | .046 | .039 | .031 | .029 | .031 | .038 | .046 | .041 |
| James-Stein+ | | | | | | | | |
| mean | -0.76 | -0.60 | -0.49 | -0.40 | -0.34 | -0.23 | -0.15 | -0.09 |
| std | .023 | .027 | .027 | .026 | .026 | .027 | .025 | .024 |
| Empirical Bayes | | | | | | | | |
| mean | -0.76 | -0.60 | -0.49 | -0.40 | -0.34 | -0.23 | -0.15 | -0.09 |
| std | .023 | .027 | .027 | .026 | .026 | .027 | .025 | .024 |

* Note that std is the standard error of the mean.

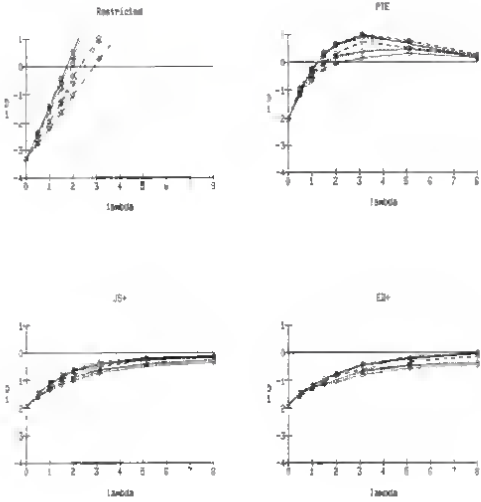



Figure 4.7 Increased Risk $D_1$ of Biased Estimates of the Cell Means over the OLS Estimate of the Cell Means, Situation 1; Table 4-7- Risk increase in $\sigma^2$ units

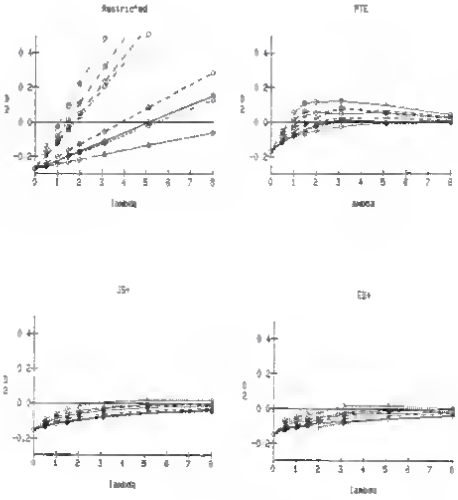



Figure 4.8 Increased Risk $D_2$ of Biased Estimates of Marginal Means over the OLS Estimate of the Marginal Means, Situation 1, Table 4.7. Risk increase in $e^2$ units.

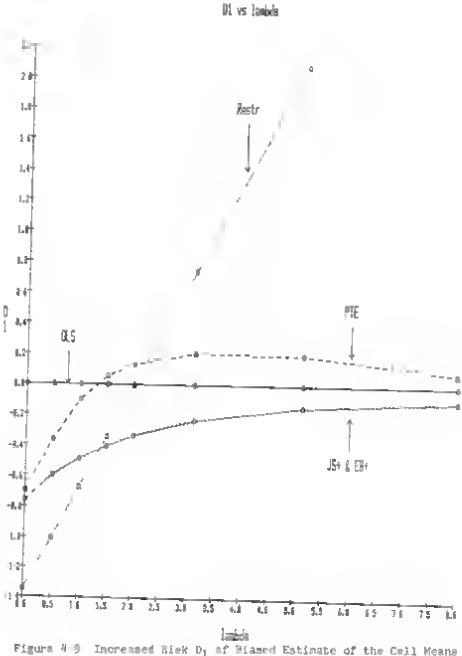



Figura 4.9 Increased Risk $D_1$ of Biased Estimate of the Cell Means over the OLS Estimate of the Cells Means, Situation 2, Table 4.7 Risk increase in $\sigma^2$ units

providing smaller confidence intervals of linear combinations of the cell means. In chapter 5 confidence intervals based on the James-Stein estimator are developed, applied in examples, and compared against other confidence intervals in simulation experiments.

CHAPTER 5
CONFIDENCE INTERVALS

In this chapter confidence intervals are developed for $\underline{l}'\underline{b}$ based upon the estimation strategies described in chapter 3. In section 2, intervals will be computed for the cell and marginal means in the factorial experiment examples given in chapter 4, using each of the described confidence intervals. Using Monte Carlo experimentation, the coverage probabilities and expected half-widths for the various confidence intervals are compared in section 3.

## 5.1 Motivating Interval Estimation Strategies

### 5.1.1 OLS Confidence Interval

The OLS point estimator of $\underline{l}'\underline{\beta}$ is $\underline{l}'\underline{b}$. The corresponding $(1-\alpha)100\%$ confidence interval for $\underline{l}'\underline{\beta}$ is

$$\underline{l}'\underline{b} \pm t(df_e,\alpha/2)\,(\underline{l}'S^{-1}\underline{l}\,s^2)^{1/2} \qquad (5.1.1)$$

Note that $t(df_e,\alpha/2)$ is the $(\alpha/2)100$ percentile of the central t distribution with $df_e$ degrees of freedom. Also, $df_e$ is the degrees of freedom associated with $s^2$, the MSE. Coverage probability for the OLS $(1-\alpha)100\%$ confidence interval is exactly $(1-\alpha)100\%$. Other interval estimation strategies will be described that achieve smaller expected half-width than this OLS confidence interval. Unfortunately, the

coverage probabilities for these interval estimation strategies are not exactly $(1-\alpha)100\%$.

5.1.2 Restricted Least Squares Confidence Interval

If the estimator of $\underline{\beta}$ is constrained to $\{\underline{\beta} \mid R\underline{\beta}=\underline{0}\}$ the resulting point estimate of $\underline{l}'\underline{\beta}$ will be $\underline{l}'(I-MR)\underline{b}$. The corresponding $(1-\alpha)100\%$ confidence interval for $\underline{l}'\underline{\beta}$ is

$$\underline{l}'(I-MR)\underline{b} \pm t(df_e,\alpha/2)\ (\underline{l}'(I-MR)S^{-1}\underline{l}\ s^2)^{1/2} \qquad (5.1.2)$$

If the constraint on $R\underline{\beta}$ holds, this restricted least squares confidence interval will have exactly $(1-\alpha)100\%$ coverage probability. When the constraint does not hold, the restricted least squares confidence interval may not maintain the desired coverage probability. The half-width for the restricted least squares interval is never larger than that of the OLS confidence interval. For some choices of $\underline{l}$, the restricted least squares confidence interval will be smaller than the OLS interval.

5.1.3 Preliminary Test Confidence Interval

By testing the hypothesis $H_0: R\underline{\beta}=\underline{0}$ to determine whether the OLS estimator or the restricted least squares estimator will be used to estimate $\underline{\beta}$ a preliminary test estimator is being used. Furthermore, if the estimate of $\underline{l}'\underline{\beta}$ takes on one of two different forms depending on the test of hypothesis, then a preliminary test estimate of $\underline{l}'\underline{\beta}$ is being employed. Similarly, a preliminary test confidence interval takes on one of two different forms, depending on the outcome of the test of hypothesis. Specifically, the following interval estimation strategy is

a preliminary test confidence interval

1. Test $H_0$ $R\underline{\beta}=\underline{0}$ using the appropriate F statistic
2. If F is greater than a critical value, c, then use [illegible] the confidence interval for $\underline{l}'\underline{\beta}$
3. If F is less than or equal to c then use (5.1.2) as [illegible] confidence interval for $\underline{l}'\underline{\beta}$.

This preliminary test confidence interval may be written using [illegible] functions as

$$\underline{l}'(\bar{\underline{\beta}} + M\,(1-1_{(0,c)}(F))\,R\underline{b}) \pm$$

$$t(df_e,\alpha/2)\,(\underline{l}'(\,(I-MR)S^{-1} + M(1-I_{(0,c)}(F))RS^{-1})\underline{l}\;s^2\,)^{1/2} \quad (5.1.3)$$

where $\bar{\underline{\beta}} = (I-MR)\underline{b}$ as in chapter 3.

The confidence interval described in (5.1.3) represents a rather severe compromise between the OLS confidence interval and the restricted least squares confidence interval. Next a compromise between the two extremes will be described that is motivated by the James-Stein compromise estimator in chapter 3.

5.1.4 James-Stein Confidence Interval

Hwang and Casella (1982) and Hwang and Chen (1986) developed and studied a confidence interval strategy in which the James-Stein point estimate is the center. Their technique was to use the James-Stein estimator in place of the OLS estimator in (5.1.1). The interval is therefore just as wide as the OLS interval. In this work an attempt will be made to obtain a reduction in interval half width. The motivation is

that the applied researcher can realize the value of shrinkage estimators if they result in confidence intervals of reduced width

A compromise confidence interval that uses a shrinkage technique in the spirit of the James-Stein estimator in chapter 3, and strives to attain (1-a)100% coverage is

$$\underline{l}'(\ (1-c/F)\underline{b} + (c/F)\underline{\tilde{\theta}}\ ) \pm$$

$$t(df_e, a/2)\ (\ \underline{l}'(\ (1-c/F)S^{-1} + (c/F)(I-MR)S^{-1}\ )\underline{l}\ \hat{\sigma}^2\ )^{1/2} \quad (5.1.4)$$

where c is the shrinkage constant described in chapter 3.

As motivation for this confidence interval, recall the model

$$y=X\underline{\theta} + \underline{\varepsilon}\ ,\quad \underline{\varepsilon} \sim N(\underline{0}, I\sigma^2) \qquad (5.1.5)$$

and the restriction

$$R\underline{\theta}=\underline{0}. \qquad (5.1.6)$$

Transforming via the method of Cook et al (1973), the model and restriction become

$$y=Z\underline{\theta} + \underline{\varepsilon} \text{ and } (I_J\ \ 0)\underline{\theta}=\underline{0}\ , \qquad (5.1.7)$$

where $Z'Z = I$. The transformation matrix is W, so that $W\underline{\theta} = \underline{\theta}$. Also, important properties of W are

$$W\ (X'X)^{-1}\ W' = I,$$

$$\text{and } W^{-1'}R'(RS^{-1}R')^{-1}RW^{-1} = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}. \qquad (5.1.8)$$

Notice that the OLS estimator of $\underline{\theta}$ is $\underline{w} = Z'\underline{y}$ , and $\underline{w}$ is $N(\underline{\theta}, I\sigma^2)$. For notation, let

$$\underline{w} = \begin{pmatrix} \underline{w}_J \\ \underline{w}_{K-J} \end{pmatrix}. \qquad (5.1.10)$$

The estimate of $\underline{\theta}$ restricted to $\{ \underline{\theta}: (I_J \; 0)\underline{\theta}=\underline{0} \}$ is

$$\bar{\underline{\theta}} = \begin{pmatrix} \underline{0}_J \\ \underline{w}_{K-J} \end{pmatrix}. \qquad (5.1.11)$$

The James-Stein compromise between $\underline{w}$ and $\bar{\underline{\theta}}$ is

$$\hat{\underline{\theta}}^S = \begin{pmatrix} \hat{\underline{\theta}}^S_J \\ \underline{w}_{K-J} \end{pmatrix} \qquad (5.1.11)$$

where $\hat{\underline{\theta}}^S_J = (1-c/F)\,\underline{w}_J + (c/F)\,\underline{0}$ . Morris (1983) describes the James-Stein estimator of a quantity such as $\underline{\theta}_J$ to be an estimate of the mean of the posterior distribution of $\underline{\theta}_J$ given $\underline{w}_J$. The posterior mean is written $(1-B)\underline{w}_J$ , and $c/F$ would be interpreted as an estimate of $B$. The posterior variance would be $(1-B)I\sigma^2$.

The proposed compromise between $S^{-1}s^2$ and $(I-MR)S^{-1}s^2$ that appears in the half-width of (5.1.4) can be related to $(1-B)I\sigma^2$, substituting $c/F$ as an estimate of $B$ and $s^2$ as an estimate of $\sigma^2$. In terms of $\underline{\theta}_J$, the estimate of posterior variance would be $(1-c/F)I_J s^2$. Using this estimated posterior variance as an estimate of the variance of $\hat{\underline{\theta}}^S$ , the resulting estimated variance of $\hat{\underline{\theta}}^S$ is

$$\hat{\mathrm{Var}}(\hat{\underline{\theta}}^S) = \begin{pmatrix} (1-c/F)I_J & 0 \\ 0 & I_{K-J} \end{pmatrix} s^2 \qquad (5.1.12)$$

The corresponding James-Stein estimator of $\underline{\beta}$ is

$$\hat{\underline{\beta}}^S = W^{-1}\hat{\underline{\theta}}^S \qquad (5.1.14)$$

An estimate of the variance of $\hat{\underline{\beta}}^J$ would be

$$\widetilde{\mathrm{Var}}(\hat{\underline{\beta}}^J) = W^{-1}\,\widetilde{\mathrm{Var}}(\hat{\underline{\theta}}^S)\,W^{-1} \qquad (5.1.15)$$

$$= W^{-1}\left[\,(1-c/F)I + (c/F)(I - \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix})\,\right] W^{-1\prime}\, s^2 \qquad (5.1.16)$$

$$= (1-c/F)W^{-1}W^{-1\prime} + (c/F)(W^{-1}W^{-1\prime} - W^{-1}\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}W^{-1})\,s^2 \qquad (5.1.17)$$

Using properties of W see in (5.1.8),

$$\widetilde{\mathrm{Var}}(\hat{\underline{\beta}}^S) = [(1-c/F)S^{-1} + (c/F)(I+WN)S^{-1}]s^2 \qquad (5.1.17)$$

The 1-α 100% confidence interval that is given in (5.1.4) can be seen to be

$$\underline{l}'\hat{\underline{\beta}}^S \pm t(df_e, \alpha/2)\,(\,\underline{l}'\,\widetilde{\mathrm{Var}}(\hat{\underline{\beta}}^S)\,\underline{l}\,)^{1/2} \qquad (5.1.18)$$

For a positive part version of this confidence interval, $\min(c_1/F,1)$ is used in place of $c/F$. The constant $c_1$ is the same as that described in chapter 3.

## 5.2 Example: Nematode Study

As the compromise estimators were applied in chapter 4 to the factorial experiment, these interval estimation strategies can be applied to the factorial experiment. Specifically, $(1-\alpha)100\%$ confidence intervals for linear combinations of the cell means can be computed using these described strategies.

In the first example of chapter 4, cell means and marginal means were estimated using the various estimators of chapter 3. Confidence intervals for the cell and marginal means are listed in Table 5.3 based on each of the strategies introduced in this chapter. The calculations required to obtain confidence intervals for the marginal mean of nematocide 1 are now shown.

The marginal mean for nematocide 1 may be written

$$\mu_{1.} = \underline{l}_1' \underline{\mu} \text{ , where } \underline{l}_1' = (1/3\ 1/3\ 1/3\ 0\ 0\ 0\ 0\ 0\ 0) \text{ .} \qquad (5.2.1)$$

The OLS 95% confidence interval for $\mu_1$ is

$$\bar{\mu}_{1.} \pm t(72, .025)(\underline{l}_1' S^{-1} \underline{l}_1 . s^2)^{1/2} . \qquad (5.2.2)$$

From a table of critical value for the central t distribution, it can be found that $t(72, .025) = 2$. The quadratic form

$$\underline{l}_1' S^{-1} \underline{l}_1 . = (1/n_{11} + 1/n_{12} + 1/n_{13})/9 \text{ ,} \qquad (5.2.3)$$

$$= (1/10 + 1/7 + 1/5)/9 = .0492$$

Referring back to Table 4.3 to get $\hat{\mu}_{1.}$, and Table 4.3 to get $s^2$, the 95% confidence interval is

$$6.209 \pm 2(.0492 \times 4.204)^{1/2}$$

which is $6.209 \pm .910$

The 95% restricted least squares confidence interval for $\mu_{1.}$ is

$$\bar{\mu}_{1.} \pm t(72, .025)(\underline{1}_1'(I-MR)S^{-1}\,\underline{1}_1 s^2)^{1/2} \qquad (5.2.4)$$

Program 1 in Appendix 8 lists $(I-MR)S^{-1}$ as part of its output. In Table 5.1 the elements of $(I-MR)S^{-1}$ are listed. Using Table 5.1, the quadratic form

$$\underline{1}_1'(I-MR)S^{-1}\,\underline{1}_1 = .0465 \qquad (5.2.5)$$

Substituting the specific values into (5.2.4), the 95% restricted confidence interval is

$$6.201 \pm 2(.0465 \times 4.204)^{1/2}$$

which is $6.201 \pm .884$

[illegible] I-MR)S [1] for Nematode Study

| | COL1 | COL2 | COL3 | COL4 | COL5 |
|---|---|---|---|---|---|
| ROW1 | 0.082a911 | 0.0141077 | 0.0148669 | 0.00900483 | -0.0[illegible] |
| ROW2 | 0.0141077 | 0.110611 | 0.0169297 | -0.00816453 | 0.0100[illegible] |
| ROW3 | 0.0148669 | 0.0169297 | 0.146565 | -0.00657931 | -0.0045164[illegible] |
| ROW4 | 0.00900483 | -0.00816453 | -0.00657931 | 0.0886508 | 0.0106137 |
| ROW5 | -0.00479766 | 0.0100798 | -0.00451647 | 0.0106137 | 0.0752921 |
| ROW6 | -0.00403846 | -0.0053425 | 0.0155564 | 0.0121989 | 0.0108949 |
| ROW7 | 0.00920451 | -0.00675963 | -0.00894553 | 0.0112095 | -0.00475466 |
| ROW8 | -0.00459798 | 0.0114848 | -0.0068827 | -0.00595989 | 0.0101228 |
| ROW9 | -0.00383878 | -0.0039376 | 0.0131902 | -0.00437466 | -0.00447348 |

| | COL6 | COL7 | COL8 | COL9 |
|---|---|---|---|---|
| | -0.00403846 | 0.00920451 | -0.00459798 | -0.00383878 |
| | -0.0053425 | -0.00675963 | 0.0114848 | -0.0039376 |
| | 0.0155564 | -0.00894553 | -0.0068827 | 0.0131902 |
| | 0.0121989 | 0.0112095 | -0.00595989 | -0.00437466 |
| | 0.0108949 | -0.00475466 | 0.0101228 | -0.00447348 |
| | 0.110052 | -0.00694057 | -0.00824461 | 0.0126543 |
| | -0.00694057 | 0.080707 | 0.00996187 | 0.00777597 |
| | -0.00824461 | 0.00996187 | 0.0808256 | 0.00767715 |
| | 0.0126543 | 0.00777597 | 0.00767715 | 0.0704557 |

The preliminary test confidence interval with pretest $\alpha = .25$ will in this case coincide with the restricted least squares confidence interval, since the intersection p-value was greater than .25.

The 95% James-Stein confidence interval for $\mu_{1.}$ may be written

$$\hat{\mu}_{1.}^{S+} \pm t(72, .025)(\underline{i}_1' D \underline{i}_1 s^2)^{1/2}, \qquad (5.2.6)$$

where $D = \min(1, c_1/F)(I-MR)S^{-1} + (I-\min(1, c_1/F))S^{-1}$. The choice of $c_1$ will be that same one used in the estimation example of chapter 4. The third program in Appendix B lists D as part of its output. In Table 5-2, the elements of D are reported for this example. Using the numbers in Table 5.2, the quadratic form can be evaluated:

$$\underline{i}_1' D \underline{i}_1 = .0480 \qquad (5.2.7)$$

Substituting the specific values into (5.2.6), the 95% James-Stein positive part confidence interval is

$$6.204 \pm 2(.0480 \times 4.204)^{1/2},$$

$$\text{or } 6.204 \pm .898$$

Using similar computations, the confidence intervals for all the cell means, and all the marginal means were computed, and are listed in Table 5.3.

| | COL1 | COL2 | [illegible] | COL4 | COL5 |
|---|---|---|---|---|---|
| ROW1 | 0.0617222 | 0.0311986 | 0.0328775 | 0.0193138 | -0.0106098 |
| ROW2 | 0.0311986 | 0.0715453 | 0.0374394 | -0.0180555 | 0.0222812 |
| ROW3 | 0.0328775 | 0.0374394 | 0.0818298 | -0.0145499 | -0.00998799 |
| ROW4 | 0.0193138 | -0.0180555 | -0.0145499 | 0.061441 | 0.0234717 |
| ROW5 | -0.0106098 | 0.0222812 | -0.00998799 | 0.061441 | 0.0563727 |
| ROW6 | -0.00893088 | -0.0118147 | 0.0344024 | 0.0269774 | 0.0240936 |
| ROW7 | 0.0203554 | -0.0149486 | -0.0197827 | 0.0247893 | -0.0105147 |
| ROW8 | -0.0101682 | 0.025398 | -0.0152208 | -0.01318 | 0.0223862 |
| ROW9 | -0.00848929 | -0.00870783 | 0.0291696 | -0.00967437 | -0.00989291 |

| COL6 | COL7 | COL8 | COL9 |
|---|---|---|---|
| -0.00893088 | 0.0203554 | -0.0101682 | -0.00848929 |
| -0.0118147 | -0.0149486 | 0.025398 | -0.00870783 |
| 0.0344024 | -0.0197827 | -0.0152208 | 0.0291696 |
| 0.0269774 | 0.0247893 | -0.01318 | -0.00967437 |
| 0.0240936 | -0.0105147 | 0.0223862 | -0.00989291 |
| 0.0703106 | -0.0153488 | -0.0182326 | 0.0275845 |
| -0.0153488 | 0.0573393 | 0.0220303 | 0.0171962 |
| -0.0182326 | 0.0220303 | 0.0575965 | 0.0169777 |
| 0.0275845 | 0.0171962 | 0.0169777 | 0.0548551 |

Table 5.3 Confidence Intervals for Nemmtode Study

| [illegible] | OLS | Restricted and PTE (α=.25) | James-Stein |
|---|---|---|---|
| [illegible] | 7.289 ± 1.297 | 7.109 ± 1.019 | 7.209 ± 1.179 |
| [illegible] | 5.179 ± 1.550 | 5.666 ± 1.087 | 5.397 ± 1.364 |
| [illegible] | 6.151 ± 1.834 | 5.828 ± 1.173 | 6.005 ± 1.570 |
| [illegible] | 8.102 ± 1.367 | 7.976 ± 1.016 | 8.045 ± 1.221 |
| [illegible] | 7.087 ± 1.236 | 6.533 ± .974 | 6.836 ± 1.125 |
| [illegible] | 5.562 ± 1.550 | 6.695 ± 1.087 | 6.129 ± 1.360 |
| [illegible] | 9.556 ± 1.297 | 9.349 ± .982 | 9.688 ± 1.165 |
| [illegible] | 8.137 ± 1.297 | 8.406 ± .984 | 8.255 ± 1.166 |
| [illegible] | 9.036 ± 1.184 | 8.568 ± .960 | 8.824 ± 1.088 |

Marginal Means

| [illegible] | | | |
|---|---|---|---|
| [illegible] | 6.209 ± .810 | 6.207 ± .884 | 6.207 ± .898 |
| [illegible] | 6.950 ± .803 | 7.068 ± .794 | 7.003 ± .799 |
| [illegible] | 8.910 ± .728 | 8.941 ± .726 | 8.924 ± .727 |
| [illegible] | | | |
| [illegible] | 8.319 ± .762 | 8.311 ± .762 | 8.314 ± .762 |
| [illegible] | 6.801 ± .790 | 6.868 ± .780 | 6.831 ± .785 |
| [illegible] | 6.950 ± .892 | 7.031 ± .854 | 6.986 ± .875 |

## 5.3 Monte Carlo Comparison

The OLS confidence interval given in (5.1.1) has exactly a $(1-\alpha)100\%$ coverage probability. The coverage probabilities for the other confidence intervals described in section 5.1 may vary, depending on the size and pattern of $R\underline{\beta}$. In this section, results of a simulation comparison of the various confidence interval strategies are given. Specifically the OLS, Restricted, Preliminary Test ($\alpha$= .25) and James Stein + 95% confidence intervals for cell means and marginal means are compared. The restriction under consideration is the additivity restriction.

### 5.3.1 Simulation Description

The same situations, interaction patterns and amounts as were used in the estimator comparison of section 4.3 are used in this confidence interval comparison. The data generation technique is also the same as the technique in section 4.3.

For each simulation-pattern-amount combination, confidence intervals are computed for the nine cell means and the three marginal means of A. These confidence intervals are computed using each of the four confidence interval strategies under study. For each pattern-amount-interval strategy combination, coverage probabilities and expected half widths of the various confidence intervals are estimated by averaging over the N=200 simulations.

As a reminder, the number of observations per cell in the two situations being simulated are given in Table 5.4.

Table 5.4 Cell Sizes Considered in the Simulations

Situation 1

| | | B 1 | B 2 | B 3 |
|---|---|---|---|---|
| A | 1 | 10 | 5 | 1 |
| | 2 | 2 | 1 | 1 |
| | 3 | 1 | 5 | 1 |

Situation 2

| | | B 1 | B 2 | B 3 |
|---|---|---|---|---|
| A | 1 | 3 | 3 | 3 |
| | 2 | 3 | 3 | 3 |
| | 3 | 3 | 3 | 3 |

In order to allow discussion of the various cells in the 3×3 factorial layout, the following numbering of cells is used:

Table 5.5 Numbering of Cells in 3×3 Layout.

| | | B 1 | B 2 | B 3 |
|---|---|---|---|---|
| A | 1 | 1 | 2 | 3 |
| | 2 | 4 | 5 | 6 |
| | 3 | 7 | 8 | 9 |

5.3.2 Simulation Results

Unbalanced cell sample sizes Simulations of the 95% confidence intervals for the nine cell means and three marginal means of A in situation 1 were performed. Estimated coverage probability and average

half width values were obtained with eight different patterns and eight different amounts of interaction. The patterns of interaction were the same as those used in section 4.3. The amounts of interaction were $\lambda$ = 0, .05, .1, .15, .2, .3, .125, .5, .12 and .5. The total number of coordinates (estimated coverage probability, average half width) estimated for confidence intervals in situation 1 was

$$(4 \text{ strategies})\times(9 \text{ cells} + 3 \text{ marginals})\times(8 \text{ amounts})\times(8 \text{ patterns})$$
$$= 3072.$$

The large number of coordinates made it prohibitive to include them in a table. Therefore, summary graphs are used to demonstrate the findings of the simulation.

Results for each cell and marginal mean are displayed in Figures 5.1 through 5.12, respectively. Notice that all the results of simulating cell one are shown in Figure 5.1. Separate graphs in Figure 5.1 contain the 64 coordinates observed for each of the four confidence interval strategies.

Imbedded in each figure are many different patterns and amounts of interaction. The desired performance for a confidence interval is small half width and coverage probability that is close to the prescribed level, irrespective of interaction pattern or amount. The four graphs in each figure allow visual comparison of the four strategies based on this criterion.

The following observations may be made upon review of Figures 5.1 through 5.12:

1. The 95% OLS confidence interval maintains 95% coverage, and always has the largest average half width

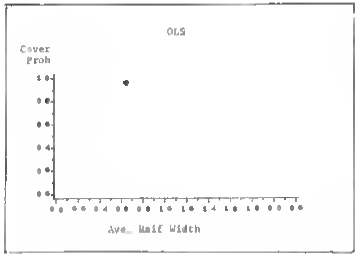
2. The 95% restricted confidence interval achieves the smallest average half width. As the interaction amounts increase, the coverage decreases. In several cases, coverage approaches 0%

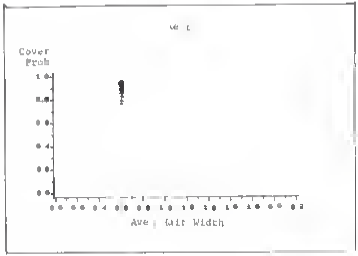
3. The 95% preliminary test ($\alpha$=.25) confidence interval offers a compromise between the average half widths of the OLS and the restricted confidence intervals. For those cells in which the average half widths of the OLS and the restricted confidence interval were most different, the coverage probability was substantially debased. Specifically, for Figures 5.3, 5.5, 5.6 and 5.7 coverage probability got as low as .6. The coverage probability was most effected for moderate amounts of interaction ($1 < \lambda < 3$). This is the range in which the F test for interaction is not very powerful, and yet the consequence of choosing the restricted confidence interval over the OLS confidence interval is substantial in terms of coverage reduction. The coverage probability was also effected noticeably in Figures 5.10 and 5.12 for the marginal means A=1 and A=3. The reason that the coverage probability is not more severely effected for the marginal means is that in this 3×3 case, the coverage reduction of the restricted confidence interval is small for moderate amounts of interaction

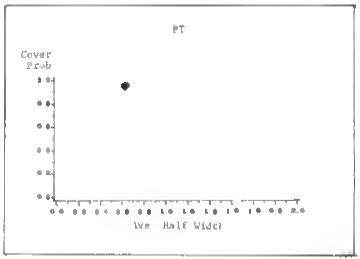
4. The 95% James Stein + confidence interval offered a compromise in average half width between those of the OLS and the

Figure 5 1 Estimated Coverage Probabilities and Average Half Widths for Call 1 in Situation 1, Table 5.4 All patterns and amounts of interaction are plotted in the same graph

CELL 1


OLS
Cover Prob
Ave. Half Width


Cover Prob
Ave. Half Width


PT
Cover Prob
Ave. Half Width

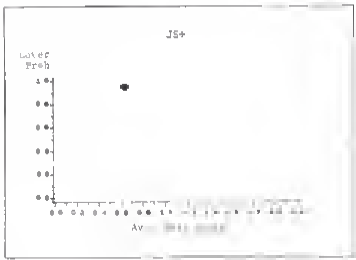
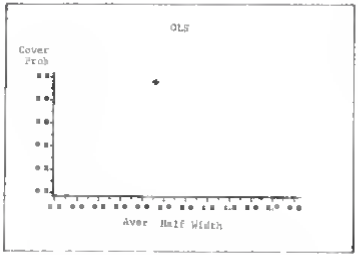
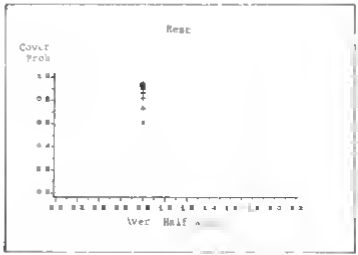
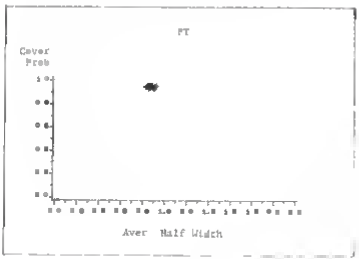

JS+
Cover Prob

Figure 5.2 Estimated Coverage Probabilities and Average Half Widths for Call 2 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph

CELL 2


OLS
Cover Prob
Aver Half Width


Rest
Cover Prob
Aver Half


FT
Cover Prob
Aver Half Width

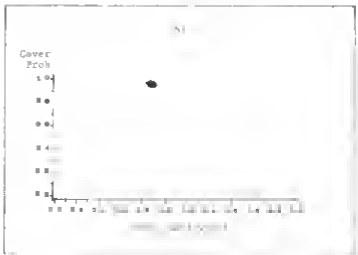
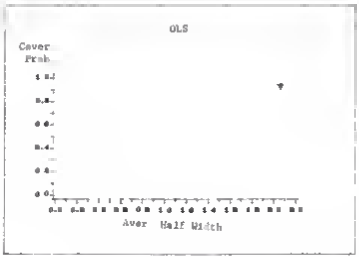
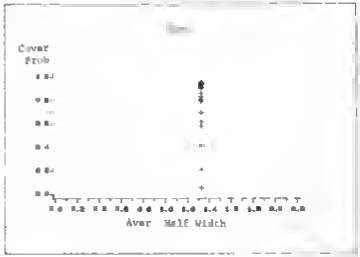
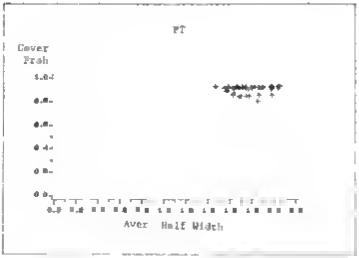

Cover Prob

Figure 5.3 Estimated Coverage Probabilities and Average Half Widths for Cell 3 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.

CELL 3


OLS
Cover
Prob
Aver Half Width


Cover
Prob
Aver Half Width


PT
Cover
Prob
Aver Half Width

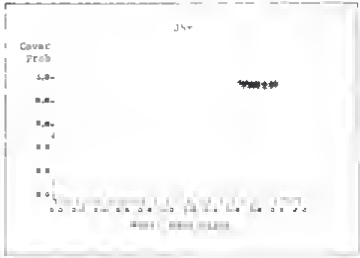
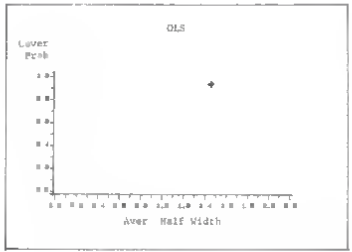
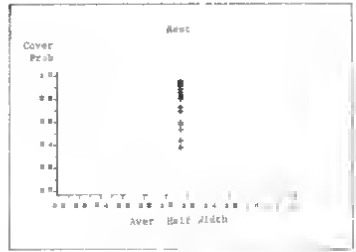
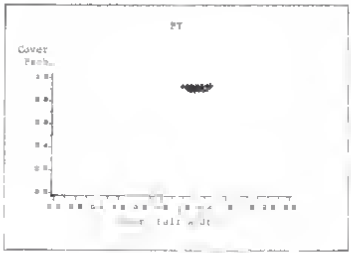

Cover
Prob

Figure 5.4 Estimated Coverage Probabilities and Average Half Widths for Cell 4 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.

CELL -


OLS
Cover
Prob
Aver Half Width


Best
Cover
Prob
Aver Half Width


PT
Cover

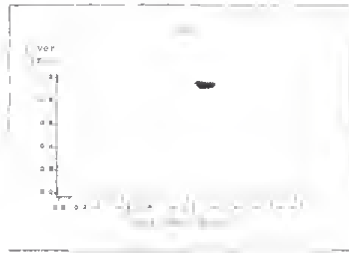
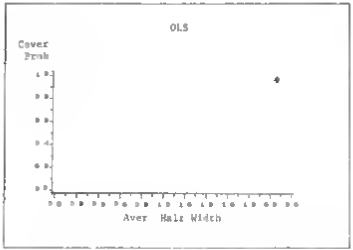
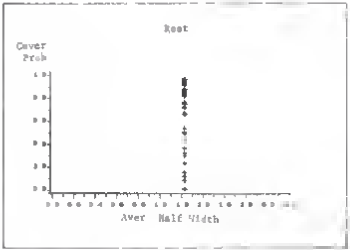
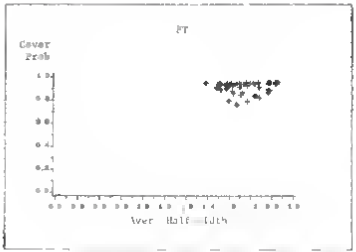


Figure 5.5 Estimated Coverage Probabilities and Average Half Widths for Cell 5 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.

CELL 5


OLS
Cover
Prob
Aver Half Width


Root
Cover
Prob
Aver Half Width


PT
Cover
Prob
Aver Half Width

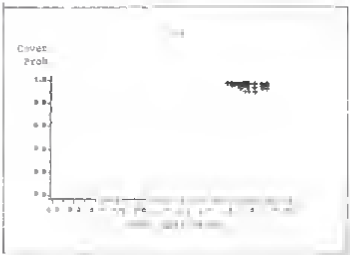
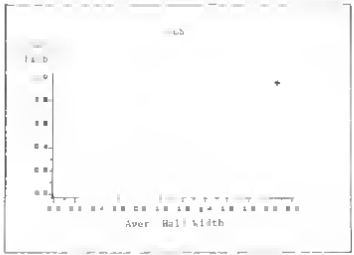
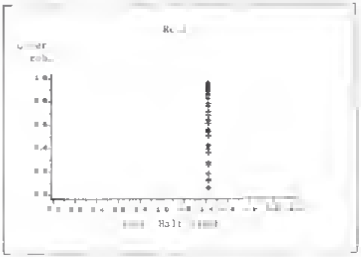
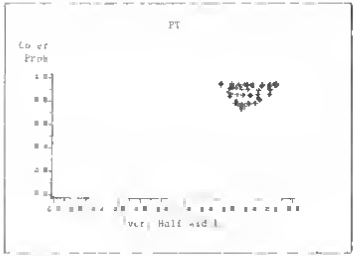

Cover
Prob

Figure 5.6 Estimated Coverage Probabilities and Average Half Widths for Cell 6 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.


Aver Half Width




PT
Cover
Prob

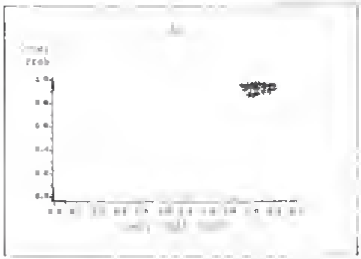
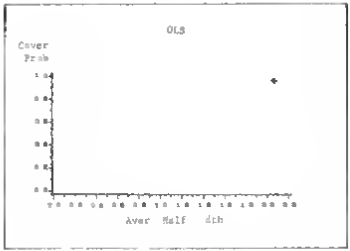
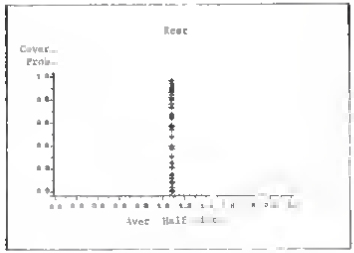
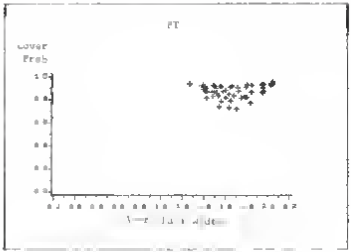

Prob

Figure 5.7 Estimated Coverage Probabilities and Average Half Widths for Cell 7 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.


OLS
Cover
Prob
Aver Half Wdth


Rear
Cover
Prob


PT
Cover
Prob

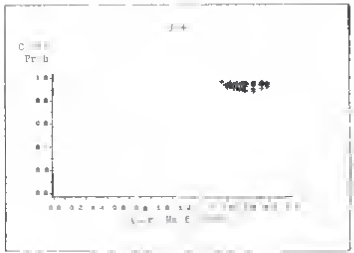
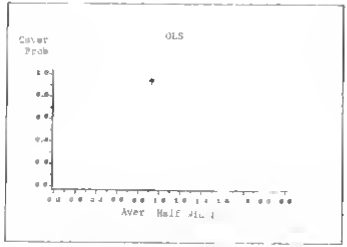
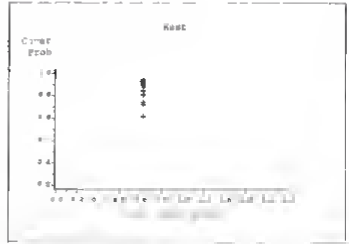
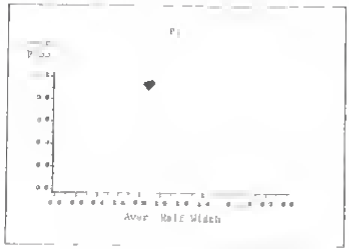

Prob

Figure 5.8 Estimated Coverage Probabilities and Average Half Widths for Call 8 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.

CELL 4


OLS
Cover
Prob
Aver Half Width


Kest
Cover
Prob


Aver Half Width

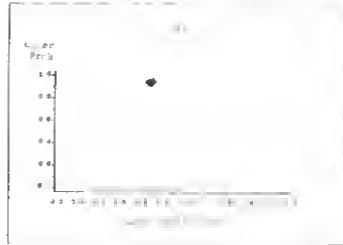


Figure 5.9 Estimated Coverage Probabilities and Average Half Widths for Cell 9 in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.

CELL 9

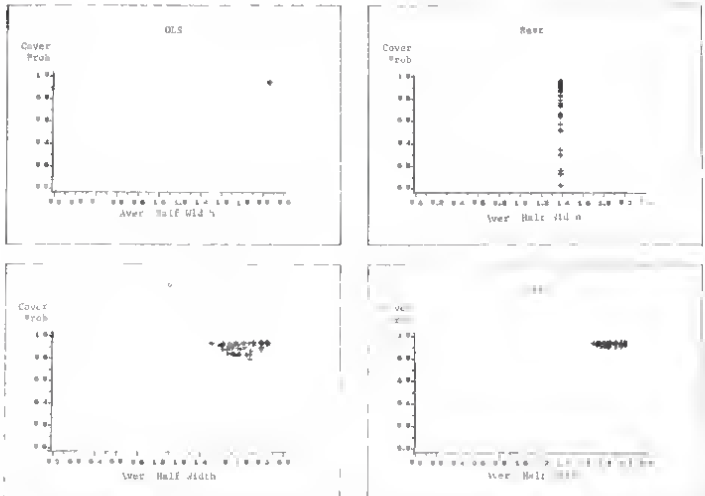
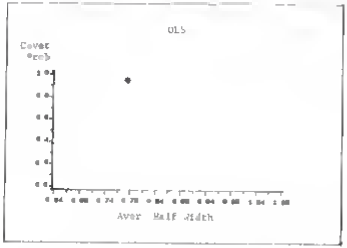
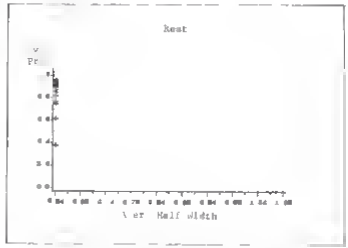
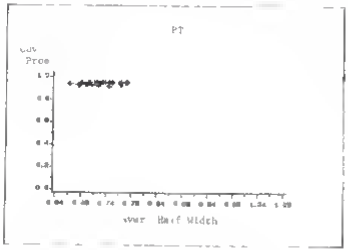

OLS
Cover
Prob
Aver Half Wid
Cover
Prob
Aver Half Wid
Cover
Prob
Aver Half Width
Aver Half

Figure 5.10 Estimated Coverage Probabilities and Average Half Widths for the A=1 Marginal Mean in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.

MARGINAL FOR A=1



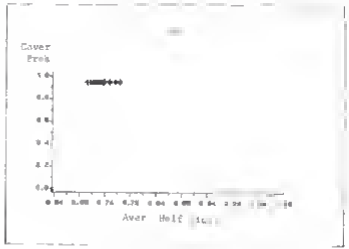
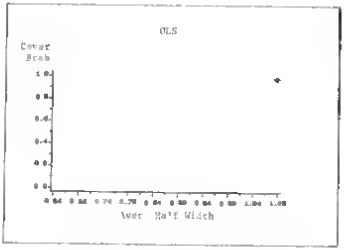
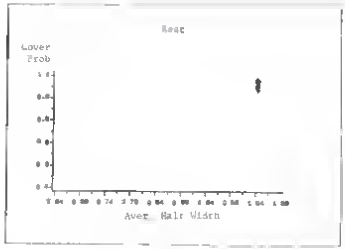
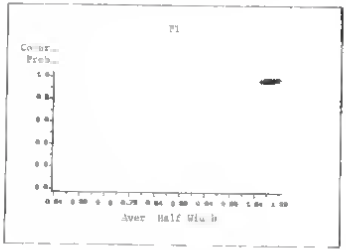



Figure 5.11 Estimated Coverage Probabilities and Average Half Widths for the A=2 Marginal Mean in Situation 1, Table 5.4 All patterns and amounts of interaction are plotted in the same graph.

MARGINAL FOR A=2



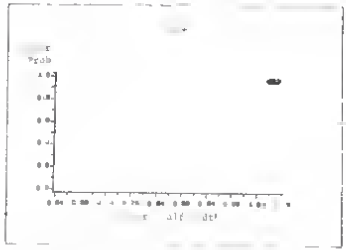
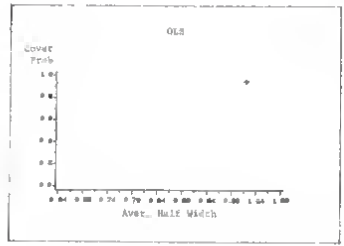
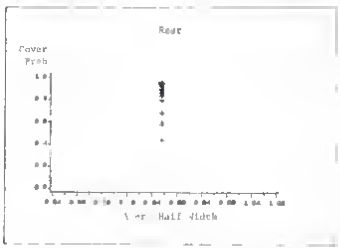
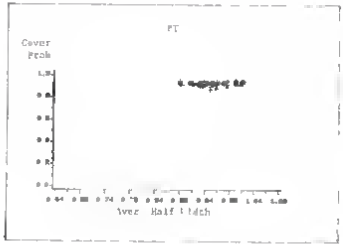
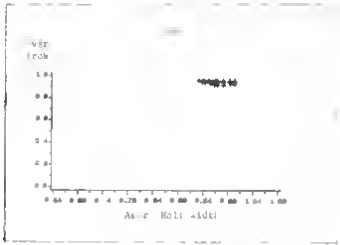



Figure 5.12 Estimated Coverage Probabilities and Average Half Widths for the A=3 Marginal Mean in Situation 1, Table 5.4. All patterns and amounts of interaction are plotted in the same graph.


OLS
Cover
Prob
Aver. Half Width


Rear
Cover
Prob


PT
Cover
Prob
Aver. Half Width



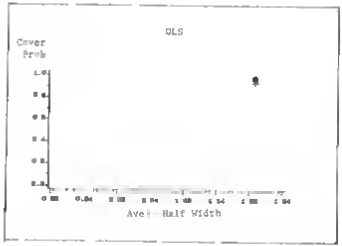
estimated confidence intervals. Coverage probability was reduced in the same cells that the preliminary confidence interval's coverage was debased. Coverage for the James Stein confidence interval got as low as .8 in some cases. In every cell, and every marginal mean simulated, the James Stein + confidence interval achieved better coverage than did the preliminary test confidence interval

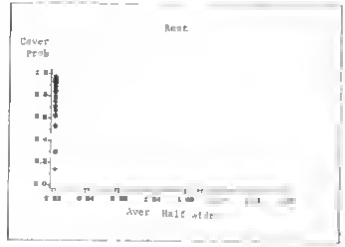
Balanced cell sample sizes. Simulation of the 95% confidence intervals for the nine cell means in situation 2 was performed. Performance coordinates were obtained for one pattern of interaction, and for $\lambda = 0$, .5, 1, 1.5, 2, 3.125, 5.12 and 8. Since the cell sample sizes are equal, all nine cell means confidence intervals should have the same performance characteristics, for a particular interval estimation strategy, and fixed $\lambda$. Therefore, in Figure 5.13 all nine cell means are included in each graph. From Figure 5.13 it appears that the James Stein + confidence intervals are not as severely effected by interaction as were the preliminary test confidence intervals.

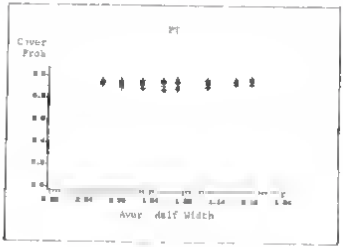
In Figure 5.14 the performances of the confidence intervals for cell mean 6 are compared for the unbalanced situation, with interaction pattern 1. In this figure, the advantage of the James-Stein confidence interval over the preliminary test interval is clearly displayed. Performances of the confidence intervals for the A=3 marginal mean are compared in Figure 5.15. Again the James-Stein confidence interval is less severely affected by increasing interaction.

Figure 5.13 Estimated Coverage Probabilities and Average Half Widths for Situation 2, Table 5.4. All nine calls are plotted in the same graph. An arbitrary pattern of interaction was used throughout.

ALL CELLS

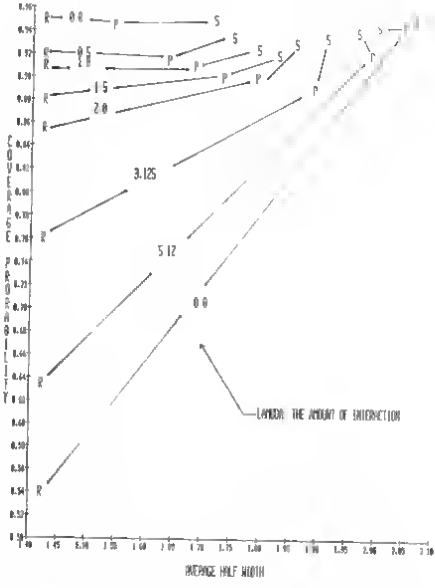

OLS
Cover Prob
Aver Half Width


Rent
Cover Prob
Aver Half Width


PT
Cover Prob
Aver Half Width

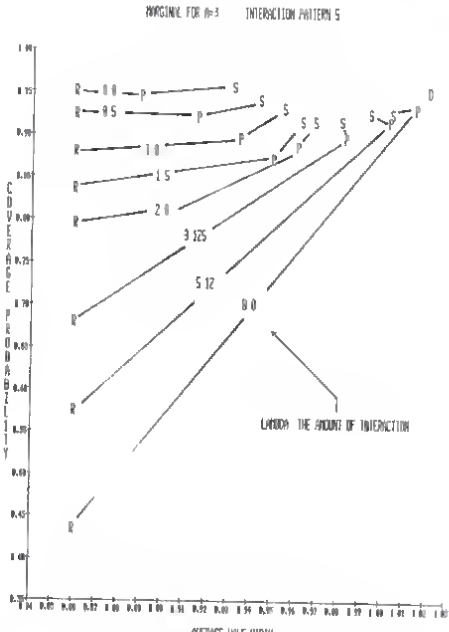
JS+
Cover Prob
Aver Half Width

CELL 6 INTERACTION PATTERN 1




Legend: O = OLS, R = Restricted, P = PT, S = James Stein +

Figure 5.14 Estimated Coverage Probabilities and Average Half Widths for the Various Confidence Intervals of Cell Mean 6 in the Unbalanced Situation




Legend: O = OLS, R = Restricted, P = PT, S = James Stein +

Figure 5.15 Estimated Coverage Probabilities and Average Half Widths for the Various Confidence Intervals of A=3 Marginal Mean in the Unbalanced Situation.

## 3.3 Conclusions

The strategy of interva[illegible]st [illegible] [illegible] on a [illegible]ict[illegible] (additivity) allows a reduction in expected half width from the exp[illegible] half widths of the corresponding OLS confidence interval. As the size of nonadditivity (interaction) increases, the coverage probability for the restricted confidence interval decreases. For this reason, the restricted confidence interval strategy is unacceptable.

The preliminary test (for interaction) interval estimation strategy allows a reduction in expected half width from that of the OLS confidence interval, as does the James-Stein + interval estimation strategy. Unfortunately, neither strategy guarantees (1-α)100% coverage probability. The James-Stein + strategy seems to hold its coverage probability a little closer to (1-α)100% than does the preliminary test strategy.

# CHAPTER 6
# CONCLUSION AND AREAS OF FUTURE RESEARCH

In this chapter, the conclusions that can be drawn from this study are stated. In addition, areas that merit further research are described.

## 6.1 Weighted Squares of Means vs Method of Fitting Constant

In chapter 2, two methods of testing for the significance of a main effect in the two-way factorial experiment were described. In the case that the cell sample sizes are unequal, the two methods can lead to conflicting conclusions about main effects. The weighted squares of means method was shown to be an unbiased testing method. The method of fitting constants method was shown to be more powerful than the weighted squares of means method, at the expense possible bias. The bias of the method of fitting constants test is shown to be most severe when the sample cell sizes are extremely unequal and interaction exists between the main effects.

As noted in chapter 2, Bancroft (1968) recommends that a preliminary test for significance of interaction be used to guide in the choice of main effect testing method. Because the power of the interaction significance test is debased by unequal sample sizes, it is expected that such a pretest test will have unacceptable bias also.

Consideration of the pretest test strategy motivated the research on estimation and confidence intervals that is presented in chapters 3

through 5. Pretest estimators [illegible] noted [illegible] shrinkage estimators by Sclove, Morris and Radhakrishan (1972) and many others. In chapter 4, application of shrinkage estimators to the two-way factorial experiment was made in an effort to determine just how the shrinkage estimators might be beneficial. In chapter 5, the concept of the pretest confidence interval was introduced. A confidence interval centered about a shrinkage estimator was developed as a competitor to the pretest confidence interval. Ultimately, study of the shrinkage strategies may motivate a compromise test of main effect, that will be a competitor to the pretest test proposed by Bancroft (1968).

## 6.2 Compromise Estimators in Factorial Experiments

In chapter 3, description and development are given for the preliminary test estimator, and its competitors--the shrinkage estimators. In chapter 4, the estimation strategies are applied to the factorial experiment. It is shown by example that computation of a James-Stein estimator of the cell means or marginal means is a relatively simple task. An empirical Bayes estimator requires more complicated computation, but can be programmed for the computer. Based on the examples that are considered, the James-Stein estimator and the empirical Bayes estimator result in very similar cell and marginal means estimates.

A simulation experiment was used to compare the performance of the preliminary test, James-Stein and empirical Bayes estimators of cell and marginal means. As might have been expected, the shrinkage estimators were found to be closer to the true cell and marginal means, on the

average. That is, the risks as defined in section 3.2 were smaller for the shrinkage estimators than for the preliminary test estimator. The differences in risk for cell mean estimation were large enough in the unbalanced situation to be of practical importance in this author's opinion.

In terms of risk for the marginal means, the differences are not necessarily large enough to warrant shrinkage estimation over preliminary test estimation. In situations of extreme cell sample size unbalance, the use of shrinkage estimators for the marginal means may provide practically significant improvement; although the cases studied in this work do not document that phenomenon. Future research in this area will attempt to identify cell sample size patterns that allow substantial improvements by using shrinkage estimators.

In the cases studied in this work, the differences between risks of the James-Stein estimators and the empirical Bayes estimator were not large enough to be of practical interest. In this author's opinion, the empirical Bayes estimator that is proposed in chapter 4 is probably not worth the extra computational complexity that it requires.

## 6.3 Confidence intervals

In chapter 5, several confidence interval strategies are described. The merits of each are now discussed.

### 6.3.1 OLS Confidence Interval

The OLS confidence interval strategy guarantees $(1-\alpha)100\%$ coverage probability for linear functions of cell means in the factorial experiment. The coverage probability is unaffected by additivity or

nonadditivity [illegible] researchers, the exact coverage probability is so attractive that all of the other methods described in chapter 5 are unacceptable. It should be noted that the exact coverage probability hinges on the assumption of normally distributed experimental error. There are very few experimental situations in which the normality assumption is actually true. Therefore, the (1-α)100% coverage probability is an approximation in practical application.

### 6.3.2 Restricted Confidence Intervals

Assuming additivity in a factorial experiment can afford substantially narrower confidence intervals for linear functions of cell means compared to the OLS strategy. The coverage probability can be severely degraded when the additivity assumption is incorrect. There are conceivably situations in which the physical mechanisms involved in the experiment prohibit the possibility of interaction. In such cases, the restricted confidence intervals would be preferrable to the OLS confidence intervals. More often though, the presence of interaction is unknown, and the assumption of additivity is fraught with danger.

### 6.3.3 Preliminary Test Confidence Interval

The preliminary test for interaction is often used in practice to facilitate the choice between OLS and restricted confidence intervals. The preliminary test confidence interval that is developed in chapter 5 accurately reflects such a decision strategy. The preliminary test confidence interval's performance is affected by the size of the interaction test and the amount and pattern in which interaction occurs. The reduction in average half width that the PT confidence interval affords does come at the expense of reduced coverage

probability [illegible] should be done [illegible] the assessment [illegible] a lower bound of coverage probability given the size of the interaction test, and a particular linear function of the cell means that an interval is being computed to capture

5.3.4 James-Stein Confidence Interval

For many experimenters, the reduction in risk that is achieved using a shrinkage estimator is useless, unless associated confidence intervals of linear combinations of the cell means are smaller than the OLS counterpart. In chapter 5, an interval estimation strategy was proposed that centers itself on the James-Stein point estimate. A compromise variance structure was developed by estimating the posterior variance of interaction. One criticism of this compromise variance structure is that it should incorporate an estimate of the variance of the compressless estimate of interaction, rather than the estimate of posterior variance of interaction (see Morris, 1983b, p. 51). The result of such a flaw is a narrower half width at the expense of some coverage probability loss.

From a practical point of view, the James-Stein confidence interval should be thought of as a straight-forward, linear compromise between OLS and restricted variance structures, using the same shrinkage constant as was used to compromise between the OLS and restricted point estimators.

The James-Stein confidence interval is narrower on the average than the OLS confidence interval. Although the coverage probability is not guaranteed, it was observed to be less affected by the presence of interaction than the PT confidence interval. Because of its obvious

relation to the James-Stein point estimation strategy, the James-Stein confidence interval that is proposed does merit consideration for practical application. Specifically, the researcher that has chosen the James-Stein estimation strategy, based on its superiority to preliminary test estimation, would naturally use the James-Stein confidence intervals to obtain approximate interval estimates.

For future research in the area of compromise confidence interval strategies, the empirical Bayes confidence intervals described by Morris (1983a) could perhaps be applied to the factorial experiment setting. Because of the added complexity of such a confidence interval, large improvements in performance over other confidence intervals would have to be documented in order to justify a recommendation for use in practice.

Another idea for future study is the study of compromise confidence interval performance in a simultaneous interval estimation strategy. It is noted by Lee (1983) that compromise estimators do not provide improvement over OLS estimation in estimating an individual component of the mean vector. The value of compromise estimators is realized in the estimation of mean vectors of dimension 3 or greater. It seems a feasible parallel to imagine that the value of compromise confidence intervals will be realized in the simultaneous interval estimation of 3 or more independent linear combinations of the mean vector.

There is not enough evidence in the work on compromise confidence intervals presented here to warrant a recommendation for their use in practice. The developed compromise confidence interval should only be considered as a proposal which can be weighed for merit in future work.

## 6.4 Recommendations

When the cell sample sizes are unequal, do not use the method of fitting constants to test the significance of a main effect.

Use a shrinkage estimate of cell means rather than a preliminary test estimate when the cell sizes are unequal.

## APPENDIX A
## GENERATING REPRESENTATIVE PATTERNS OF INTERACTION FOR SIMULATION OF A TWO-WAY FACTORIAL EXPERIMENT

In chapter 4 and chapter 5, computer simulation of data corresponding to a two-way factorial experiment was used as a basis for performance comparison of various estimators, and confidence intervals. In the case that the number of observation per cell in a factorial experiment are not all equal, the pattern in which interaction occurs in the data can have an effect on the performance of the various estimators and confidence intervals that have been introduced. It was therefore necessary that several different patterns of interaction be considered in the simulations. This appendix describes the generation of eight interaction patterns that are representative of the population of all possible interaction patterns. Specifically, two of the patterns used in the simulation were chosen because they were known cause $R_1(\underline{\mu},\underline{\mu})$ to be maximum and minimum for any fixed choice of $\lambda_i$. Two other interaction patterns were chosen because they maximized and minimized $R_2(\underline{\mu},\underline{\mu})$ for any fixed $\lambda_i$. Finally, four patterns of interaction were chosen randomly from the population of all possible interaction patterns.

## [illegible] $R_1(\tilde{\underline{\mu}},\underline{\mu})$

[illegible]

$$R_1(\tilde{\underline{\mu}},\underline{\mu}) = E(\ (\tilde{\underline{\mu}}-\underline{\mu})'(\tilde{\underline{\mu}}-\underline{\mu})\ )$$

$$= E(\ (\tilde{\underline{\mu}}-\underline{\mu})'(I-MR)\hat{\underline{\mu}}_1(I-MR)\underline{\mu}\ ) + R_1((M\underline{0},MR\underline{\mu}) \quad \text{from lemma 3.4}$$

$$= E(\ (\tilde{\underline{\mu}}-\underline{\mu})'(I-MR)'(I-MR)(\hat{\underline{\mu}}-\underline{\mu})\ ) + E(\ (\underline{0}-R\underline{\mu})'M'M(\underline{0}-R\underline{\mu})\ )$$

$$= tr(\ (I-MR)'(I-MR)S^{-1}\sigma^2) \ + \ \underline{\mu}'M'M\underline{\mu}$$

$$= \sigma^2 tr(S^{-1})-\sigma^2 tr(MRS^{-1})+\underline{\mu}'R'M'MR\underline{\mu} \qquad \text{(A.1)}$$

Following the reparameterization of Cook et.al (1973), let $\underline{\theta} = W\underline{\mu}$ such that (1.1-3) becomes

$$\underline{y} = Z\ \underline{\theta} + \underline{\varepsilon} \qquad \text{(A.2)}$$

and $R\underline{\mu} = \underline{0}$ becomes

$$[I_J \quad 0]\underline{\theta} = \underline{0}, \qquad \text{(A.3)}$$

where $Z'Z = I_K$, $J = (a-1)(b-1)$, and $k = ab$. Take $W = QP$ where $P$ is a nonsingular matrix such that $P(X'X)^{-1}P' = I_K$ and $Q$ is an orthogonal matrix such that

$$Q(P^{-1})'R'(RS^{-1}R\ )^{-1}RP^{-1}Q' = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{(A.4)}$$

In terms of $\underline{u}$, $[I_J \; 0]\underline{e}$, and $\lambda$ is $\lambda = \underline{d}_J'\underline{e}_J/2\sigma^2$. Also,

$$\underline{u}'R'M'M\underline{u} = \underline{d}_J A\underline{d}_J, \qquad \text{A-5}$$

where $A = [I_J \; 0] \, (W^{-1})' \, W^{-1} \, [I_J \; 0]'$.

From Graybill (1969, p. 309), we know that

$$d_J \, \underline{e}_J' \, \underline{e}_J \leq \underline{e}_J' \, A \, \underline{e}_J \leq d_1 \, \underline{e}_J' \, \underline{e}_J \, , \qquad \text{(A-6)}$$

where $d_1$ and $d_J$ are the largest and smallest roots of A, respectively.

Take $\underline{e}_{(1)}$ and $\underline{e}_{(J)}$ to be eigenvectors corresponding to $d_1$ and $d_J$, respectively. Also, assume $\underline{e}_{(1)}'\underline{e}_{(1)} = \underline{e}_{(J)}'\underline{e}_{(J)} = 1$.
By properties of eigenvectors,

$$\underline{e}_{(1)}' \, A \, \underline{e}_{(1)} = d_1$$

$$\text{and} \qquad \text{(A-7)}$$

$$\underline{e}_{(J)}' \, A \, \underline{e}_{(J)} = d_J .$$

Let $\underline{f}_1 = W^{-1}\begin{pmatrix} \underline{e}_{(1)} \\ \underline{0} \end{pmatrix}$, and $\underline{f}_J = W^{-1}\begin{pmatrix} \underline{e}_{(J)} \\ \underline{0} \end{pmatrix}$. Notice that

$$\underline{f}_1'R'M'MR\underline{f}_1 = d_1$$

$$\text{and} \qquad \text{(A-8)}$$

$$\underline{f}_J'R'M'MR\underline{f}_J = d_J$$

Also, $\underline{f}_1'R'(RS^{-1}R'[illegible])\underline{f}_1 = 1$, and $\underline{f}_J'R'(RS^{-1}R'[illegible])\underline{f}_J = [illegible]$ Therefore, if $\underline{\mu} = \underline{f}_1$ then $\lambda = (2\sigma^2)^{-1}$ and $\underline{\mu}'R'M'MR\underline{\mu} = d_1$. For $\lambda = (2\sigma^2)^{-1}$, $d_1$ is the largest value that $\underline{\mu}'R'M'MR\underline{\mu}$ can attain. Also, if $\underline{\mu} = \underline{f}_J$ then $\lambda = (2\sigma^2)^{-1}$, and $\underline{\mu}'R'M'MR\underline{\mu} = d_J$. For $\lambda = (2\sigma^2)^{-1}$, $d_J$ is the smallest value that $\underline{\mu}'R'MR\underline{\mu}$ can attain.

So $\underline{f}_1$ and $\underline{f}_J$ represent patterns of interaction that allow (A.1) to take on its maximum and minimum value for fixed $\lambda$, respectively.

## A.2 Patterns of Interaction that Maximize (Minimize) $R_2(\bar{\underline{\mu}},\underline{\mu})$

The risk of the restricted estimator is

$$R_2(\bar{\underline{\mu}},\underline{\mu}) = tr(S^{-1}D^*) - tr(MRS^{-1}D^*) + \underline{\mu}'R'M^{-1}D^*MR\underline{\mu}, \qquad (A.9)$$

where $D^*$ is as in (3.2.3). Let $\underline{g}_1$ and $\underline{g}_J$ be such that $\underline{g}_1 = W^{-1}\begin{pmatrix}\underline{e}^*_{(1)} \\ \underline{0}\end{pmatrix}$, and $\underline{g}_J = W^{-1}\begin{pmatrix}\underline{e}^*_{(J)} \\ \underline{0}\end{pmatrix}$ where $\underline{e}^*_{(1)}$ and $\underline{e}^*_{(J)}$ are eigenvectors corresponding to the largest and smallest roots of $A^*$. The matrix $A^* = [I_J 0]\, W^{-1} D^* W^{-1} [I_J 0]$. The eigenvectors can be chosen so that $\underline{e}^{*\prime}_{(1)}\underline{e}^*_{(1)} = \underline{e}^{*\prime}_{(J)}\underline{e}^*_{(J)} = 1$. The vectors $\underline{g}_1$ and $\underline{g}_J$ give the patterns of interaction such that (A.9) is minimized, and maximized for fixed $\lambda$, respectively.

## A.3 Random Patterns of Interaction

The following algorithm was used to generate random patterns of interaction:

1. Generate a vector $\underline{z}$ where $\underline{z}$ is a real [illegible] $N_d(\underline{0}, I)$ random vector. Standardize $\underline{z}$ so that

$$\underline{z} = \frac{\underline{z}}{|\underline{z}|}$$

2. The vector $\underline{r} = W^{-1}(\underline{z}' \; \underline{0}')'$ is then a random pattern of interaction, such that $\lambda = 1$.

In Table A.1, the patterns of interaction that were used in the simulation of the unbalanced situation (see Table 4.5 or Table 5.4) are given. Since these eight patterns span the realm of all possible interaction patterns well, the simulation results for risk estimation (in Chapter 4) and confidence interval performance (in Chapter 5) are indicative of the results one might get using different interaction patterns.

Table A.1 Patterns of Interaction for the Simulation of the Unbalanced Case. [illegible]

| Cell (i,j) | Interaction Pattern 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1,1 | -0.037685 | -0.092701 | -0.059569 | -0.011815 | -0.090312 | 0.011611 | 0.084163 | 0.000859 |
| 1,2 | 0.003710 | 0.169542 | 0.208425 | 0.030233 | 0.169045 | -0.015985 | -0.050471 | 0.043520 |
| 1,3 | 0.358297 | 0.079295 | -0.048438 | -0.033019 | 0.057896 | -0.036182 | -0.589270 | -0.226131 |
| 2,1 | -0.067178 | 0.307044 | 0.285100 | -0.054490 | 0.083772 | -0.263257 | -0.194647 | -0.096247 |
| 2,2 | -0.179209 | -0.670583 | -0.354904 | 0.488042 | 0.024073 | -0.073861 | -0.104042 | -0.477392 |
| 2,3 | 0.313565 | 0.055495 | -0.215297 | 0.597023 | -0.201616 | 0.600395 | 0.433335 | 0.569886 |
| 3,1 | 0.511203 | 0.312917 | 0.425488 | 0.227125 | 0.735576 | 0.410427 | -0.452332 | 0.183900 |
| 3,2 | 0.032132 | -0.035425 | -0.137844 | 0.067376 | -0.175353 | 0.030757 | 0.071280 | 0.051359 |
| 3,3 | -0.671861 | -0.135791 | 0.263735 | -0.564004 | 0.143720 | -0.564213 | 0.094934 | -0.443695 |
| Type of Pattern | maximizes $R_1(\underline{\mu},\underline{\mu})$ | random | minimizes $R_1(\underline{\mu},\underline{\mu})$ | random | maximizes $R_2(\underline{\mu},\underline{\mu})$ | minimizes $R_2(\underline{\mu},\underline{\mu})$ | random | random |

## APPENDIX H
## COMPUTER PROGRAMS FOR COMPUTING BIASED ESTIMATES OF CELL MEANS

In this appendix simple computer programs are given for computing various cell means estimators. The programs are all written for SAS, PROC MATRIX. These programs are set up so that given information, such as the number of observations per cell, the size of the factorial experiment, the average responses in each cell, and the F statistic testing $H_0$: no interaction, the compromised estimators can be computed. The following programs are included

Program 1 - Computes restricted estimator.

Program 2 — Computes conditions on the shrinkage constant that will insure minimaxity of the James-Stein estimators in estimating cell means

Program 3 - Computes James-Stein positive part estimator.

Program 4 - Computes modified Empirical Bayes positive part estimator

Following each program is an example output, from a run of the program.

The following instructions will enable a user to alter these programs so that they apply to any two-way factorial experiment.

1. [illegible] the factorial experiment [illegible] programs through the variables A and B. The variable [illegible] assigned by the user to be the number of levels of factor A in the experiment

$$A = (\# \text{ of levels of factor A})$$

Similarly, B is assigned by the user to be the number of levels of factor B in the experiment.

$$B = (\# \text{ of levels of factor B})$$

2. The number of observations in each factor level combination is entered by the user through the vector NIJ, where

$$NIJ = \begin{matrix} n_{11} & n_{12} & \cdots & n_{1b} \\ n_{21} & n_{22} & \cdots & n_{2b} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ n_{a1} & n_{a2} & \cdots & n_{ab} \end{matrix}$$

3. The sample averages for each factor level combination (ordinary least squares estimators) are entered by the user through the vector OLS, where

$$\bar{y}_{11} \quad \bar{y}_{22} \quad \cdots \quad \bar{y}_{ab}$$

- The value of the F statistic, testing for in [illegible] (1) entered into the program by the user;

  F = (observed interaction F stat.);

- For the programs that compute shrinkage estimators, the value of the shrinkage constant may be specified by typing in a specific number. If a constant is not specified by the user, the constant will take on a default value. Program 2 can be used to determine a value of the shrinkage constant for the James-Stein pos. part estimator so that the resulting estimator will be minimax with respect to $R_1(\quad , \quad)$ as in (3.2.1). If the default value is desired, specify

  C1 = .,  or  C = .;

  for the James-Stein positive part estimator or the Empirical Bayes positive part, respectively.

## Program 1: Computes Restricted Estimator

```
PROC MATRIX
A=3 B=3 COMMENT THE SIZE OF THE EXPT. IS GIVEN
NIJ=10  7  5
     9 11  7
    10 10 12;COMMENT THE NUMBERS OF OBS  PER CELL IS GIVEN;
NIJ=NIJ',
OLS=7.285 5.179 6.151 8.102 7.087 5.662
    9.556 8.137 9.038;COMMENT CELL AVERAGES ARE GIVEN
OLS=OLS',
PRINT A B NIJ OLS,
COMMENT THE RESTRICTION MATRIX IS DEVELOPED;
R1=J(A,1,1) R2=J(B,1,1),
DO K=1 TO A,R1(K,1)=K END;
DO K=1 TO B R2(K,1)=K END;
O1=ORPOL(R1) O2=ORPOL(R2);
R=(O1(,2 A))'#(O2(,2:B))';
COMMENT THE RESTRICTED ESTIMATOR IS COMPUTED;
S=DIAG(NIJ) S_1=INV(S);
M= S_1*R'*INV(R*S_1*R')
REST=OLS - M*R*OLS,
PRINT REST;
COMMENT (I - MR) S_1 IS PRINTED FOR USE IN CONSTRUCTING CONFIDENCE
INTERVALS
D=S_1 - M*R*S_1
PRINT D
NOTE D IS (I-MR)S_1,
```

## 8.2 Example Results From Program 1

| A | COL1 |
|---|---|
| ROW1 | 3 |

| R | COL1 |
|---|---|
| ROW1 | 3 |

| NIJ | COL1 |
|---|---|
| ROW1 | 10 |
| ROW2 | 7 |
| ROW3 | 5 |
| ROW4 | 9 |
| ROW5 | 11 |
| ROW6 | 7 |
| ROW7 | 10 |
| ROW8 | 10 |
| ROW9 | 12 |

| OLS | COL1 |
|---|---|
| ROW1 | 7.289 |
| ROW2 | 5.179 |
| ROW3 | 6.151 |
| ROW4 | 8.102 |
| ROW5 | 7.087 |
| ROW6 | 5.662 |
| ROW7 | 9.556 |
| ROW8 | 8.137 |
| ROW9 | 9.036 |

| REST | COL1 |
|---|---|
| ROW1 | 7.10938 |
| ROW2 | 5.6661 |
| ROW3 | 5.8283 |
| ROW4 | 7.97603 |
| ROW5 | 6.53274 |
| ROW6 | 6.65494 |
| ROW7 | 9.4899 |
| ROW8 | 8.40571 |
| ROW9 | 8.56791 |

(The matrix D is printed also. See Table 5.1 for the D matrix of this example.)

## B.3 Program 2. Choosee Shrinkage Constant to Insure Minimaxity of James-Stein Estimator

```
PROC MATRIX;
COMMENT THIS PROGRAM WILL FIND A RANGE OF VALUES FOR C IN THE J-S
        ESTIMATOR TO INSURE MINIMAXITY OF THE J_S ESTIMATOR.
A=3;B=3; COMMENT THE SIZE OF THE EXPT  IS GIVEN
NIJ=10  7  5
    9 11  7
   10 10 12  COMMENT THE NUMBERS OF OBS  PER CELL IS GIVEN
NIJ=NIJ';
PRINT A B NIJ
COMMENT THE RESTRICTION MATRIX IS DEVELOPED.
R1=J(A,1,1) R2=J(B,1,1);
DO K=1 TO A;R1(K,1)=K;END;
DO K=1 TO B;R2(K,1)=K;END;
O1=ORPOL(R1),O2=ORPOL(R2);
R=(O1(,2:A)'@(O2(,2:B))');
COMMENT COMPUTATION OF ROOTS AND TRACE OF MRS_1;
S=DIAG(NIJ) S_1=INV(S); N=SUM(NIJ),
MRS_1=S_1*R'*INV(R*S_1*R')*R*S_1,
TR=TRACE(MRS_1);
ROOT=EIGVAL(MRS_1);
DL=ROOT(1,)
DENOM=(A-1)#(B-1)#(N-A#B+2),
C=((A-1)#(B-1)-2)#(N-A#B)#/DENOM;
NOTE THE STANDARD CHOICE OF C IS;
PRINT C
NOTE CONDITIONS FOR MINIMAXITY ARE ,
NOTE 1)TR=TRACE(MRS_1) GT 2*DL, DL=LARGEST;
NOTE ROOT OF MRS_1;
PRINT TR DL
UPBND=2#(TR#/DL-2)#(N-A#B)#/DENOM;
NOTE 2) 0 LE C LE UPBND
PRINT UPBND
NOTE IF CONDITION 2) IS NOT MET WITH THE STANDARD CHOICE,
NOTE OF C, THEN A BETTER CHOICE SHOULD BE USED.
```

B 4 Example Results From Program 2

| A | COL1 |
|---|---|
| ROW1 | 3 |

| B | COL1 |
|---|---|
| ROW1 | 3 |

| NIJ | COL1 |
|---|---|
| ROW1 | 10 |
| ROW2 | 7 |
| ROW3 | 5 |
| ROW4 | 9 |
| ROW5 | 11 |
| ROW6 | 7 |
| ROW7 | 10 |
| ROW8 | 10 |
| ROW9 | 12 |

THE STANDARD CHOICE OF C IS

| C | COL1 |
|---|---|
| ROW1 | 0.486486 |

CONDITIONS FOR MINIMAXITY ARE
1)TR=TRACE(MRS_1) GT 2*DL, DL=LARGEST
ROOT OF MRS_1

| TR | COL1 |
|---|---|
| ROW1 | 0.49806 |

| DL | COL1 |
|---|---|
| ROW1 | 0.167356 |

2) 0 LE C LE UPBND

| UPBND | COL1 |
|---|---|
| ROW1 | 0.474837 |

IF CONDITION 2) IS NOT MET WITH THE STANDARD CHOICE
OF C, THEN A BETTER CHOICE SHOULD BE USED

B-5 Program 3 : Computes James-Stein Positive Part Estimator

```
PROC MATRIX;
A=3;B=3, COMMENT THE SIZE OF THE EXPT  IS GIVEN
NIJ=10  7  5
     9 11  7
    10 10 12; COMMENT THE NUMBERS OF OBS  PER CELL IS GIVEN,
NIJ=NIJ';
OLS=7.285 5.179 6.151
    8.102 7.087 5.662
    5.556 8.137 9.036; COMMENT CELL AVERAGES ARE GIVEN;
OLS=OLS';
C1=.4746; COMMENT UNLESS A SPECIFIC VALUE FOR C1 IS GIVEN HERE, A
          DEFAULT CHOICE FOR C1 IS USED,
F=1.05;COMMENT THE F STA. FOR RD NO INTERACTION;
PRINT A B NIJ OLS F;
COMMENT THE RESTRICTION MATRIX IS DEVELOPED;
R1=J(B,1,1);R2=J(B,1,1),
DO K=1 TO A;R1(K,1)=K,END,
DO K=1 TO B;R2(K,1)=K;END
O1=ORPOL(R1);O2=ORPOL(R2)
R=(O1(,2:A))'#(O2(,2:B))',
COMMENT THE J-S POS. PART ESTIMATOR IS COMPUTED;
S=DIAG(NIJ),S_1=INV(S);N=SUM(NIJ)
M= S_1*R'*INV(R*S_1*R');
REST=OLS - M*R*OLS;
IF C1=. THEN
C1=((A-)#(B-1)-2)#(N-A#B)#/((A-1)#(B-1)#(N-A#B+2));
ADJ=C1#/F,
IF ADJ GT 1 THEN ADJ=1,
JSPLUS=(1 - ADJ)#OLS + ADJ#REST;
PRINT C1 JSPLUS;
COMMENT (1-C1/F)S_1 + (C1/F)(I-MR)S_1 IS PRINTED FOR USE IN COMPUTING
CONFIDENCE INTERVALS;
D=(1-ADJ)#S_1 + ADJ#(S_1 - M*R*S_1);
PRINT D;
NOTE D IS (1-C1/F)S_1 + (C1/F)(I-MR)S_1;
```

## 4.6 Example Results From Program 3

| A | COL1 |
|---|---|
| ROW1 | 3 |

| B | COL1 |
|---|---|
| ROW1 | 3 |

| NIJ | COL1 |
|---|---|
| ROW1 | 10 |
| ROW2 | 7 |
| ROW3 | 5 |
| ROW4 | 9 |
| ROW5 | 11 |
| ROW6 | 7 |
| ROW7 | 10 |
| ROW8 | 10 |
| ROW9 | 12 |

| OLS | COL1 |
|---|---|
| ROW1 | 7.289 |
| ROW2 | 5.179 |
| ROW3 | 6.151 |
| ROW4 | 8.102 |
| ROW5 | 7.087 |
| ROW6 | 5.662 |
| ROW7 | 9.556 |
| ROW8 | 8.137 |
| ROW9 | 9.036 |

| F | COL1 |
|---|---|
| ROW1 | 1.05 |

| C1 | COL1 |
|---|---|
| ROW1 | 0.4748 |

| JSPLUS | COL1 |
|---|---|
| ROW1 | 7.20778 |
| ROW2 | 5.39526 |
| ROW3 | 6.00508 |
| ROW4 | 8.04504 |
| ROW5 | 6.83637 |
| ROW6 | 6.12909 |
| ROW7 | 9.68849 |
| ROW8 | 8.25851 |
| ROW9 | 8.52433 |

(The matrix D is printed a[illegible] see Table 5.2 for the D matrix [illegible] example.)

B 7 Program 4 Computes Empirical Bayes Positive Part Estimator

```
PROC MATRIX;
A=3;B=3, COMMENT THE SIZE OF THE EXPT IS GIVEN;
NIJ=10 7 5
    9 11 7
    10 10 12, COMMENT THE NUMBERS OF OBS PER CELL IS GIVEN,
NIJ=NIJ';
OLS=7.289 5.179 6.151
    5.102 7.087 5.662
    8.556 8.137 9.036 COMMENT CELL AVERAGES ARE GIVEN;
OLS=OLS';
G= , COMMENT UNLESS A SPECIFIC VALUE FOR G IS GIVEN HERE, A DEFAULT
    CHOICE FOR G IS USED,
P=1.05;COMMENT THE F STAT FOR HO NO INTERACTION,
PRINT A B NIJ OLS F;
COMMENT THE RESTRICTION MATRIX IS DEVELOPED
R1=J(A,1,1);R2=J(B,1,1)
DO K=1 TO A,R1(K,1)=K END,
DO K=1 TO B,R2(K,1)=K END,
O1=ORPOL(R1);O2=ORPOL(R2),
R=(O1(,2:A))'@(O2(,2:B))'
S=DIAG(NIJ),S_1=INV(S)_N=SUM(NIJ)
P=INV(R'S_1*R'),TR=TRACE(P);RF=NROW(P),
M=S_1*R'*P,
BEST=OLS - M*R*OLS;
COMMENT THE E-B POS PART ESTIMATOR IS COMPUTED;
IF G=. THEN
G=((A-1)#(B-1)-2)#(N-A#B-2)#/((A-1)#(B-1)#(N-A#B+2)),
VHAT=((N-A#B-2)#P#/(N-A#B)-1)#(A-1)#(B-1) #/ TR,
E_B=OLS-G#M*INV(VHAT#P + I(RP) )*R*OLS;
PRINT E_B
COMMENT THE AMOUNT OF SHRINKAGE INVOLVED FOR EACH CELL MEAN IS
COMPUTED,
SHRINK=(E_B OLS)#/(BEST-OLS),
COMMENT THE ADJUSTMENT TO MAKE THIS A POSITIVE PART VERSION IS MADE,
DO INDX=1 TO A#B;
IF SHRINK(INDX,1) LT 0 THEN E_B(INDX,1)=OLS(INDX,1),
IF SHRINK(INDX,1) GT 1 THEN E_B(INDX,1)=BEST(INDX,1),
END,
SHRINK=(E_B-OLS)#/(BEST-OLS),
PRINT SHRINK
NOTE SHRINK TELLS THE AMOUNT THAT EACH CELL MEAN OLS,
NOTE EST HAS BEEN SHRUNK TOWARD THE RESTRICTED ESTIMATE
```

## 5.8 Example Results From Program 4

| A | COL1 |
|---|---|
| ROW1 | 3 |

| B | COL1 |
|---|---|
| ROW1 | 3 |

| NIJ | COL1 |
|---|---|
| ROW1 | 10 |
| ROW2 | 7 |
| ROW3 | 5 |
| ROW4 | 9 |
| ROW5 | 11 |
| ROW6 | 7 |
| ROW7 | 10 |
| ROW8 | 10 |
| ROW9 | 12 |

| OLS | COL1 |
|---|---|
| ROW1 | 7.289 |
| ROW2 | 5.179 |
| ROW3 | 6.151 |
| ROW4 | 8.102 |
| ROW5 | 7.087 |
| ROW6 | 5.662 |
| ROW7 | 9.556 |
| ROW8 | 6.137 |
| ROW9 | 9.036 |

| F | COL1 |
|---|---|
| ROW1 | 1.05 |

| S_B | COL1 |
|---|---|
| ROW1 | 7.2064 |
| ROW2 | 5.40532 |
| ROW3 | 5.99937 |
| ROW4 | 8.04315 |
| ROW5 | 6.83048 |
| ROW6 | 6.14076 |
| ROW7 | 9.69157 |
| ROW8 | 8.26075 |
| ROW9 | 8.8199 |

| SHRINK | COL1 |
|---|---|
| ROW1 | 0.458892 |
| ROW2 | 0.464522 |
| ROW3 | 0.469889 |
| ROW4 | 0.467138 |
| ROW5 | 0.462815 |
| ROW6 | 0.463493 |
| ROW7 | 0.462696 |
| ROW8 | 0.460522 |
| ROW9 | 0.461656 |

SHRINK TELLS THE AMOUNT THAT EACH CELL MEAN OLS
EST HAS BEEN SHRUNK TOWARD THE RESTRICTED ESTIMATED

## BIBLIOGRAPHY


Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance. Ann. Math. Statist. 15(1), 190-204.

Bendroff, T.A. (1968). Topics in intermediate Statistical Methods. Iowa State University Press, Ames.

Bancroft, T.A., and Han, C.-P. (1977). Inference based on conditional specification: A note and a bibliography. International Statistical Review 45, 117-127.

Baranchik, A.J. (1964). Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution. Tech. Report 51. Department of Statistics, Stanford University, Stanford CA.

Berger, J. (1976a). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. Ann. Statist. 4, 223-226.

Berger, J. (1976b). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, 256-264.

Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. 8, 716-761.

Bickel, P.J. (1984). Parametric robustness: Small biases can be worthwhile. Ann. Statist. 12, 1349-1368.

Bock, M.E., Yancey, T.A., and Judge, G.G. (1973). The statistical consequences of preliminary test estimators in regression. J. Amer. Statist. Assoc. 68, 109-116.

Casella, G. (1985). An introduction to empirical Bayes data analysis. American Statistician, 39, 83-87.

Cochran, W.G., and Cox, G.M. (1957). Experimental Designs, 2nd ed. Wiley, New York.

Cramer, E.M., and Appelbaum, M.I. (1980). Nonorthogonal analysis of variance-once again. Psychological Bulletin 87, 51-57.

Efron, B., and Morris, C. (1973). Stein estimation rule and its competitors--An empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117-130.

Ghosh, M., and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. J. Amer. Statist. Assoc. 81 (in press).

Gosslee, D.G., and Lucas, H.L. (1965). Analysis of variance of disproportionate data when interaction is present. Biometrics 21, 115-133.

Graybill, F.A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth, Belmont, CA.

Gun, A.M. (1967). Use of a preliminary test in the estimation of factorial means. Calcutta Statistical Association Bulletin, 16, 49-72.

Hocking, R.R. (1981). An analysis of methods used in ANOVA with missing cells. Proc. 6th Ann. SUGI Conf., pp. 94-99.

Hwang, J.T., and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. Ann. Statist. 10, 868-881.

Hwang, J.-T., and Chen, J. (1986). Improved confidence sets for the coefficients of a linear model with spherically symmetric errors. Ann. Statist., 14, 444-460.

James, W., and Stein, C. (1961). Estimation with quadratic loss. Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability, pp. 361-379. University of California Press, Berkeley, CA.

Johnson, N.L., and Kotz, Samuel, (1970). Continuous Univariate Distributions-2. Houghton Miffin, New York.

Judge, G.G., and Bock, M.E. (1978). The Statistical Implications of Pretest and Stein-rule Estimators in Econometrics. North-Holland, New York.

Lee, Y. (1983). Estimation of Multivariate Normal Mean and Its Application to Mixed Linear Models. Thesis. Iowa State University, Ames.

Morris, C. (1983a). Parametric empirical Bayes confidence intervals. Scientific Inference, Data Analysis and Robustness (G.E.P. Box, T. Leonard, and C. Wu, eds.), pp. 25-50. Academic Press, New York.

[illegible], F. ([illegible]). Parametric empirical Bayes inference [illegible] and [illegible]tions. J. Amer. Statist. Assoc. 76, 47-65.

Oman, S.D. (1982). Shrinking towards subspaces in multiple linear regression. Technometrics 24, 307-317.

Overall, J.E., Lee, D.M., and Hornick, C.W. (1981). Comparison of two strategies for analysis of variance in nonorthogonal designs. Psychological Bulletin 90, 367-375.

Sclove, S., Morris, C., and Radhakrishnan, R. (1972). Nonoptimality of preliminary test estimators of the mean of multivariate normal distribution. Ann. Math. Statist. 43, 1481-1490.

Searle, S.R. (1971). Linear Models. Wiley, New York.

Searle, S.R. (1979). Relationships between the estimable functions of SAS GLM output for unbalanced data and the hypotheses tested by traditional-style F-statistics. Proc. 4th Ann. SUGI Conf., pp. 196-207.

Searle, S.R. (1981). Quirks in linear model computations for unbalanced data. Proc. 6th Ann. SUGI Conf., pp. 38-45.

Searle, S.R., Speed, F.M., and Henderson, H.V. (1981). Some computational and model equivalences in analysis of variance of unequal-subclass-numbers data. The American Statistician 35, 16-33.

Speed, F.M., and Hocking, R.R. (1980). A characterization of the GLM sums of squares. Proc. 5th Ann. SUGI Conf., pp. 215-218.

Speed, F.M., Hocking, R.R., and Hackney, O.P. (1978). Methods of analysis of unbalanced data. J. Amer. Statist. Assoc. 73, 105-112.

Steel, R.G.D., and Torrie, J.H. (1960). Principles and Procedures of Statistics. McGraw-Hill, New York.

Steel, R.G.D., and Torrie, J.H. (1980). Principles and Procedures of Statistics, 2nd ed. McGraw-Hill, New York.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution. J. Royal Statist. Soc. B24, 265-296.

Stein, C. (1966). An approach to the recovery of interblock information in balanced incomplete block designs. Festschrift for J. Neyman (F.N. David, ed.), pp. 351-366. Wiley, New York.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135-1151.

[illegible] W. (19[illegible]). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Math. Statist., 42, 385-388.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different subclasses. J. Amer. Statist. Assoc. 29, 51-56.

Yates, F. (1935). Complex experiments. J. Royal Statist. Soc. Suppl. 2, 181-247.

Yates, F. (1937). The design and analysis of factorial experiments. Imp. Bur. Soil Sci. Tech. Comm., No. 35.

## BIOGRAPHICAL SKETCH

Richard was born August 23, 1956. At the age of two his family moved to central Florida, where he attended public school from grades 1 to 12. Specifically, he attended Union Park Elementary, Union Park Junior High, and Oviedo High. During the 1974/1975 school year, Richard majored in mathematics at Texas Lutheran College and tried to play college baseball. Failing as a ballplayer, he returned to central Florida and earned the Bachelor of Science degree in statistics from the University of Central Florida in 1978. In 1981 he received the Master of Statistics degree from University of Florida. In the summer of 1984, Richard left Gainesville to work at Harris Semiconductor in Melbourne, Florida.

Richard's wife is Susan, the daughter of William and June Noonan. His parents are Marvin and Virgie Lynch. Richard has three sisters, Marleen, Lydie and Joy.

Richard is a Christian and looks forward to being more involved in the Central Baptist Church, Melbourne, after graduation.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ramon C. Littell, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andrawus I. Khuri
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John S. Shonkwiler
Associate Professor of Food and Resource Economics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1986

Dean, Graduate School